



# UNIVERSIDAD TÉCNICA PARTICULAR DE LOJA

*La Universidad Católica de Loja*

## MODALIDAD PRESENCIAL

## ESCUELA DE CIENCIAS DE LA COMPUTACIÓN

**Aplicación de técnicas de minería de datos para identificar patrones de comportamientos relacionados con las acciones del estudiante con el EVA de la UTPL.**

Trabajo de fin de carrera previa la obtención del título de Ingeniero en Sistemas Informáticos y Computación.

**Autora:**

*Sarango Sedamanos Marcia Yudy*

**Directora:**

*Ing. Valdiviezo Díaz Priscila Marisela*

**Coodirectora:**

*Ing. Arias Tapia Susana Alexandra*

**LOJA-ECUADOR**

**2012**

## **CERTIFICACIÓN**

Ing. Priscila Valdiviezo

## **DIRECTORA DE TESIS**

### **CERTIFICA:**

Que la Srta. Marcia Yudy Sarango Sedamanos, autor de la tesis APLICACIÓN DE TÉCNICAS DE MINERÍA DE DATOS PARA IDENTIFICAR PATRONES DE COMPORTAMIENTOS RELACIONADOS CON LAS ACCIONES DEL ESTUDIANTE CON EL EVA DE LA UTPL, ha cumplido con los requisitos estipulados en el Reglamento General de la Universidad Técnica Particular de Loja, la misma que ha sido coordinada y revisada durante todo el proceso de desarrollo, desde su inicio hasta la culminación, por lo cual autorizo su presentación.

Loja, Diciembre del 2011

---

**Ing. Priscila Valdiviezo**

## **CESIÓN DE DERECHOS**

Yo, Marcia Yudy Sarango Sedamanos, declaro ser autor del presente trabajo y eximo expresamente a la Universidad Técnica Particular de Loja y a sus representantes legales de posibles reclamos o acciones legales.

Adicionalmente declaro conocer y aceptar la disposición del Art. 67 del Estatuto Orgánico de la Universidad Técnica Particular de Loja, que en su parte pertinente textualmente dice: “Forman parte del patrimonio de la Universidad la propiedad intelectual de investigaciones, trabajos científicos o técnicos y tesis de grado que se realicen a través o con el apoyo financiero, académico o institucional (operativo) de la Universidad”.

---

**Marcia Yudy Sarango Sedamanos**

## AUTORÍA

Las ideas, opiniones, conclusiones, recomendaciones y más contenidos expuestos en el presente informe de tesis son de absoluta responsabilidad del autor.

---

**Marcia Yudy Sarango Sedamanos**

## **DEDICATORIA**

A Dios, por ser mi luz y nunca soltarme de su mano en el caminar, por sus bendiciones pues sus planes son perfectos.

A mis amados padres, por ser mi ejemplo constantemente en el trayecto de la vida, por brindarme la oportunidad de culminar el sueño que un día tuve, siempre en cada decisión y cada paso que doy los tengo presentes.

A mis hermanos, por la compañía y el apoyo que me brindan, se que cuento y contaré con ellos en todo momento.

A mi abuelita (+) mamita, por su paciencia, amor incondicional y demostrarme que se puede.

A todas las personas con las que conté, se interesaron y estuvieron ahí con una palabra de apoyo para que siga adelante y lleve a buen término mi trabajo.

***Marcia***

## AGRADECIMIENTO

Esta tesis implica una serie de sucesos y personas para lograr un sueño que con paciencia y dedicación se puede conseguir.

Agradezco al Ing. Nelson Piedra, que en calidad de director de la escuela de Ciencias de la Computación y docente apoya al estudiante inculcando el valor investigativo, innovador y humano, para el desarrollo profesional.

A la Ing. Priscila Valdiviezo, por su dirección, exigencia y motivación durante el desarrollo de mi tesis; a la Ing. Susana Arias por su guía como codirectora.

A la Ing. Samanta Cueva, por sus indicaciones en mi informe de tesis final.

También extendiendo mi gratitud a los docentes, compañeros y amigos que compartieron sus conocimientos y experiencias conmigo durante mis años de estudio en la universidad.

Con especial cariño.

***Marcia***

## Índice de Contenidos

|  |       |
|--|-------|
| <b>CERTIFICACIÓN</b> .....   | ii    |
| <b>CESIÓN DE DERECHOS</b> .....  | iii   |
| <b>AUTORÍA</b> .....   | iv    |
| <b>DEDICATORIA</b> .....   | v     |
| <b>AGRADECIMIENTO</b> .....  | vi    |
| <b>ÍNDICE DE CONTENIDOS</b> .....  | vii   |
| <b>ÍNDICE DE TABLAS</b> .....  | xii   |
| <b>ÍNDICE DE FIGURAS</b> .....   | xv    |
| <b>RESUMEN</b> .....   | xviii |
| <b>Capítulo 1 Estado del Arte</b> .....  | 1     |
| 1.1 Introducción .....   | 2     |
| 1.2 Modelo de Usuario .....  | 3     |
| 1.3 Técnicas para el modelo de usuario .....   | 4     |
| 1.4 Modelo de Estudiante .....   | 6     |
| 1.5 Características a tener en cuenta en la construcción de un modelo del estudiante ..... | 7     |
| 1.6 Técnicas para la construcción de modelo estudiante abierto .....                       | 8     |
| 1.6.1 <i>¿Cómo construir un modelo del estudiante?</i> .....                               | 9     |
| 1.7 Estilos de aprendizaje y modelado de usuario .....                                     | 11    |
| 1.7.1 <i>Adaptabilidad y preferencias de aprendizaje</i> .....                             | 12    |
| 1.8 Ambientes de aprendizaje e interactividad .....  | 14    |
| 1.9 Características de los ambientes interactivos .....                                    | 14    |
| 1.10 Datos de usuario .....  | 15    |
| 1.10.1 <i>Datos demográficos</i> .....   | 16    |
| 1.11 Datos de uso .....  | 16    |
| 1.11.1 <i>Uso observables</i> .....  | 17    |
| 1.11.2 <i>Regularidades de uso</i> .....   | 18    |

|   |   |           |
|---|---|-----------|
| 1.12  | Tipo de interacción del usuario .....   | 19        |
| 1.13  | Minería de datos para el modelo del estudiante .....  | 20        |
| 1.14  | El Proceso de minería de datos en la enseñanza a distancia .....  | 22        |
| 1.15  | Técnicas de minería de datos .....  | 23        |
| 1.16  | Construcción del modelo de predicción .....   | 27        |
| 1.17  | Trabajos relacionados .....   | 28        |
| 1.17.1  | <i>Inferencia en el modelo de usuario .....</i>   | 28        |
| 1.17.2  | <i>La minería de datos en los sistemas de gestión de cursos: Caso de estudio de Moodle tutorial.....</i>  | 29        |
| 1.17.3  | <i>Minería de datos al estudiante para caracterizar el comportamiento similar en grupos de espacios de colaboración estructurada .....</i>  | 30        |
| 1.17.4  | <i>Aplicación de minería de datos con una herramienta de software en la evaluación del rendimiento académico de los alumnos de la carrera de Sistemas de la FACENA-UNNE .....</i> | 31        |
| <b>Capítulo 2 Análisis del EVA de la UTPL .....</b>   |   | <b>32</b> |
| 2.1   | Introducción .....  | 33        |
| 2.2   | Moodle .....  | 34        |
| 2.2.1   | <i>Herramientas de apoyo para el aprendizaje .....</i>  | 34        |
| 2.2.2   | <i>Herramientas de apoyo para el aprendizaje en el EVA de la UTPL.....</i>  | 35        |
| 2.3   | Acciones de los estudiantes en el EVA .....   | 38        |
| 2.4   | Modelo Entidad/Relación del EVA .....   | 39        |
| 2.4.1   | <i>Entidades de moodle relacionadas con las acciones de los estudiantes en el eva .....</i>   | 40        |
| <b>Capítulo 3 Minería de datos para determinar comportamientos del estudiante relacionados con las acciones en el EVA .....</b> |   | <b>43</b> |
| 3.1   | Descripción del escenario .....   | 44        |
| 3.1.1   | <i>Algoritmos para determinar los comportamientos del estudiante en el EVA.....</i>   | 45        |
| 3.1.2   | <i>Análisis de los datos .....</i>  | 46        |

|  |  |     |
|--|--|-----|
| 3.1.3  | <i>Indicadores correspondientes a las acciones del estudiante en el EVA</i>  | 47  |
| 3.1.3  | <i>Escenario de experimentación</i>  | 51  |
| 3.2  | Limpieza y transformación de datos   | 54  |
| 3.3  | Discretización   | 55  |
| 3.3.1  | <i>Discretización para los valores de los atributos de la carrera de informática</i>   | 55  |
| 3.3.2  | <i>Discretización para los valores de los atributos de la carrera de abogacía</i>  | 61  |
| 3.4  | Minería de datos   | 66  |
| 3.5  | Experimentación y resultados con los algoritmos de clasificación   | 67  |
| 3.5.1  | <i>Resultados obtenidos para la carrera de informática</i>   | 67  |
| 3.5.2  | <i>Resultados obtenidos para la carrera de abogacía</i>  | 70  |
| 3.5.3  | <i>Comparación de las matrices de confusión entre la carrera de informática y abogacía</i>   | 73  |
| 3.6  | Experimentación con algoritmo de clustering  | 75  |
| 3.6.1  | <i>Experimentación y resultados del curso de lógica de la programación</i>   | 75  |
| 3.6.2  | <i>Experimentación y resultados del curso de ética y derechos humanos</i>  | 95  |
| 3.6.3  | <i>Comparación de los grupos entre los cursos: lógica de la programación y, ética y derechos humanos obtenidos con el algoritmo de clustering SimpleKMeans</i> | 115 |
| 3.7  | Minería de datos para determinar estilos de aprendizaje  | 118 |
| 3.7.1  | <i>Introducción</i>  | 118 |
| 3.7.2  | <i>Modelo de estilos de aprendizaje</i>  | 119 |
| 3.8  | Población y muestra de estudio   | 121 |
| 3.9  | Experimentación y resultados   | 124 |
| <b>Capítulo 4 Representación de indicadores y atributos del estudiante en un modelo de datos</b> |  |     |
|  |  | 132 |
| 4.1  | Introducción   | 133 |

|       |  |     |
|-------|--|-----|
| 4.4.1 | <i>Propuesta del modelo del estudiante usando la minería de datos</i> .....  | 133 |
| 4.2   | Descripción de la metodología a seguir .....   | 134 |
| 4.3   | Proceso del diseño del modelo de datos del estudiante .....  | 136 |
| 4.3.1 | <i>Definición de las entidades</i> .....   | 136 |
| 4.3.2 | <i>Definición de los atributos</i> .....   | 137 |
| 4.3.3 | <i>Definición de las relaciones y cardinalidad</i> .....   | 139 |
| 4.4   | Modelo de datos del estudiante .....   | 141 |
| 4.4.1 | <i>Descripción de las entidades del modelo de datos del estudiante</i> .....   | 143 |
|       | Discusión .....  | 150 |
|       | Conclusiones y recomendaciones .....   | 152 |
|       | <i>Conclusiones</i> .....  | 153 |
|       | <i>Recomendaciones</i> .....   | 155 |
|       | Trabajos futuros .....   | 156 |
|       | <b>Referencias Bibliográficas</b> .....  | 157 |
|       | <b>ANEXOS</b> .....  | 163 |
|       | <b>ANEXO A Migración de la base de datos</b> .....   | 164 |
|       | ANEXO A.1 Migración de la base de datos .....  | 165 |
|       | <b>ANEXO B Sentencias SQL para preprocesamiento de los datos</b> .....   | 166 |
|       | ANEXO B.1 Consultas SQL.....   | 167 |
|       | ANEXO B.2 Consultas SQL para identificar los registros de cada las acciones en cada<br>modulo y curso .....                            | 170 |
|       | ANEXO B.3 Consultas SQL para identificar los registros de los intentos para rendir un<br>cuestionario con sus respectivos tiempos..... | 170 |
|       | <b>ANEXO C Test de estilos de aprendizaje</b> .....  | 172 |
|       | ANEXO C.1 Test de Estilos de Aprendizaje. ....   | 173 |
|       | ANEXO C.2 Test de Felder y Silverman. ....   | 173 |
|       | <b>ANEXO D Carga de los datos en Weka</b> .....  | 178 |

|   |            |
|---|------------|
| ANEXO D.1 Conexión de la base de datos MineríaAcciones en mysql con Weka..... | 179        |
| <b>ANEXO E Paper</b> .....  | <b>195</b> |

## Índice de Tablas

|   |    |
|---|----|
| Tabla 1.1: Clasificación de estilos de aprendizaje-enseñanza (Espinoza, Maza y Lopez,2009) ..           | 13 |
| Tabla 2.1: Atributos de las acciones del usuario sobre la plataforma .....                              | 38 |
| Tabla 2.2: mdl_enrol_utpl .....   | 40 |
| Tabla 2.3: mdl_user .....   | 40 |
| Tabla 2.4: mdl_log .....  | 41 |
| Tabla 2.5: mdl_twt_utpl .....   | 41 |
| Tabla 2.6: mdl_quiz_attempts .....  | 42 |
| Tabla 3.1: Muestra de las carreras con sus respectivos cursos .....                                     | 44 |
| Tabla 3.2: Algoritmos para experimentación .....  | 45 |
| Tabla 3.3: Acciones en las herramientas utilizadas en los cursos .....                                  | 50 |
| Tabla 3.4: Discretización para el nivel de participación en el curso (informática) .....                | 57 |
| Tabla 3.5: Discretización para el nivel de utilización de herramientas (informática) .....              | 60 |
| Tabla 3.6: Discretización para el nivel de participación en el curso (abogacía) .....                   | 62 |
| Tabla 3.7: Discretización para el nivel de utilización de herramientas (abogacía) .....                 | 65 |
| Tabla 3.8: Resultados obtenidos para el nivel de participación en el curso (informática) .....          | 67 |
| Tabla 3.9: Resultados obtenidos para el nivel de utilización de herramientas (informática) .....        | 68 |
| Tabla 3.10: Comparación del la razón de precisión del algoritmo para cada indicador (informática) ..... | 69 |
| Tabla 3.11: Algoritmo seleccionado para cada indicador (informática) .....                              | 70 |
| Tabla 3.12: Resultados obtenidos para el nivel de participación en el curso (abogacía) .....            | 70 |
| Tabla 3.13: Resultados obtenidos para el nivel de utilización de herramientas (abogacía) .....          | 71 |

|             |  |     |
|-------------|--|-----|
| Tabla 3.14: | Comparación del la razón de precisión del algoritmo para cada indicador (abogacía).                  | 72  |
| Tabla 3.15: | Algoritmo seleccionado para cada indicador (abogacía) .....  | 73  |
| Tabla 3.16: | Comparación de las matrices de confusión .....   | 73  |
| Tabla 3.17: | Primer experimento con el algoritmo SimpleKMeans-semilla 1000 (lógica de la programación) .....      | 76  |
| Tabla 3.18: | Segundo experimento con el algoritmo SimpleKMeans-semilla 3000 (lógica de la programación) .....     | 78  |
| Tabla 3.19: | Tercer experimento con el algoritmo SimpleKMeans-semilla 7000 (lógica de la programación) .....      | 80  |
| Tabla 3.20: | Comparativa de los experimentos realizados utilizando SimpleKMeans (lógica de la programación) ..... | 82  |
| Tabla 3.21: | Nivel de utilización de las herramientas (lógica de la programación) .....                           | 83  |
| Tabla 3.22: | Primer experimento con el algoritmo SimpleKMeans-semilla 1000 (ética y derechos humanos) .....       | 96  |
| Tabla 3.23: | Segundo experimento con el algoritmo SimpleKMeans-semilla 3000 (ética y derechos humanos) .....      | 99  |
| Tabla 3.24: | Tercer experimento con el algoritmo SimpleKMeans-semilla 7000 (ética y derechos humanos) .....       | 101 |
| Tabla 3.25: | Comparativa de los experimentos realizados utilizando SimpleKMeans (ética y derechos humanos) .....  | 102 |
| Tabla 3.26: | Nivel de utilización de las herramientas (ética y derechos humanos) .....                            | 103 |
| Tabla 3.27: | Comparación de grupos .....  | 115 |
| Tabla 3.28: | Variable estilos de aprendizaje. Bases Teóricas (Felder y Silverman) .....                           | 118 |
| Tabla 3.29: | Correspondencia entre los estilos de aprendizaje. Bases Teóricas: (Felder y Silverman) .....         | 123 |
| Tabla 3.30: | Primer experimento con el algoritmo FarthestFirst – semilla 5 .....                                  | 125 |
| Tabla 3.31: | Segundo experimento con el algoritmo FarthestFirst – semilla 20 .....                                | 126 |

|             |  |     |
|-------------|--|-----|
| Tabla 3.32: | Tercer experimento con el algoritmo FarthestFirst – semilla 100 .....                | 127 |
| Tabla 3.33: | Resultados de correspondencia- clúster 0 .....                                       | 128 |
| Tabla 3.34: | Resultados de correspondencia- clúster 1 .....                                       | 129 |
| Tabla 4.1:  | Definición de entidades .....  | 136 |
| Tabla 4.2:  | Definición de atributos .....  | 137 |
| Tabla 4.3:  | Definición de relaciones y cardinalidad .....  | 140 |
| Tabla 4.4:  | Atributos entidad mdl_user .....   | 143 |
| Tabla 4.5:  | Atributos entidad mdl_course .....   | 144 |
| Tabla 4.6:  | Atributos entidad model_user .....   | 144 |
| Tabla 4.7:  | Atributos participacion_curso .....  | 145 |
| Tabla 4.8:  | Atributos utilizacion_herramientas .....   | 146 |
| Tabla 4.9:  | Atributos entidad resultados_estilos_aprendizaje .....                               | 147 |
| Tabla 4.10: | Atributos entidad acciones_en_objetos .....  | 148 |
| Tabla 4.11: | Atributos entidad acciones .....   | 149 |
| Tabla 4.12: | Atributos entidad objetos .....  | 149 |
| Tabla B.1:  | Consultas para identificar el curso y su periodo correspondiente (informática) ..... | 168 |
| Tabla B.2:  | Consultas para identificar el curso y su periodo correspondiente (abogacia) .....    | 169 |

## Índice de Figuras

|  |    |
|--|----|
| Figura 1.1: Esquema general de un modelo de usuario (Vélez, 2009) .....  | 5  |
| Figura 1.2: Aspectos a tener en cuenta en la construcción de un modelo del estudiante .....  | 7  |
| Figura 1.3: Fases del KDD (Hernández, Ramírez y Ferri, 2008) .....   | 21 |
| Figura 1.4: Pasos de los métodos de aplicación de minería de datos en la enseñanza a distancia (Romero, Ventura, y García, 2007) ..... | 22 |
| Figura 1.5: Relación adquisición de conocimiento – fenómeno causal - actividad .....   | 27 |
| Figura 2.1: E/R que contienen información de las acciones de los estudiantes .....   | 39 |
| Figura 3.1: Escenario para el nivel de participación en el curso y el nivel de utilización de las herramientas por carrera .....       | 51 |
| Figura 3.2: Escenario para el nivel de utilización de las herramientas por curso .....   | 52 |
| Figura 3.3: Escenario para encontrar el estilo de aprendizaje dominante.....   | 53 |
| Figura 3.4: Consulta SQL para extraer atributos de la participación en el curso (informática) .....                                    | 57 |
| Figura 3.5: Asignación de la clase para el indicador del nivel de participación (informática) .....                                    | 58 |
| Figura 3.6: Consulta SQL para extraer atributos de la utilización de herramientas (informática)...                                     | 59 |
| Figura 3.7: Asignación de la clase para el indicador de la utilización de la herramienta (informática) .....                           | 61 |
| Figura 3.8: Consulta SQL para extraer atributos de la participación en el curso (abogacía).....  | 61 |
| Figura 3.9: Asignación de la clase para el indicador del nivel de participación (abogacía) .....                                       | 63 |
| Figura 3.10: Consulta SQL para extraer atributos de la utilización de herramientas (abogacía) .....                                    | 64 |
| Figura 3.11: Asignación de la clase para el indicador de la utilización de la herramienta (abogacía) .....                             | 66 |
| Figura 3.12: Número de veces que contesta un foro según número de accesos al curso (lógica de la programación) .....                   | 85 |
| Figura 3.13: Número de veces que descarga un recurso según el número de accesos al curso (lógica de la programación) .....             | 86 |

|  |     |
|--|-----|
| Figura 3.14: Número de veces que envía o sube una tarea según número de acceso al curso (lógica de la programación).....   | 87  |
| Figura 3.15: Número de veces que revisa el historial de mensajes según el numero de mensajes que ha escrito (lógica de la programación) .....                    | 88  |
| Figura 3.16: Número de mensajes twitter según número de acceso al curso (lógica de la programación) .....  | 89  |
| Figura 3.17: Número de intentos por rendir una evaluación según el tiempo utilizado (evaluación a distancia primer bimestre), (lógica de la programación) .....  | 90  |
| Figura 3.18: Número de intentos por rendir una evaluación según el tiempo utilizado (evaluación a distancia segundo bimestre), (lógica de la programación) ..... | 91  |
| Figura 3.19: Número de veces que contesta un foro según número de accesos al curso (ética y derechos humanos) .....  | 104 |
| Figura 3.20: Número de veces que descarga un recurso según el número de accesos al curso (ética y derechos humanos) .....  | 105 |
| Figura 3.21: Número de veces que envía o sube una tarea según número de acceso al curso (ética y derechos humanos) .....   | 106 |
| Figura 3.22: Número de veces que revisa el historial de mensajes según el numero de mensajes que ha escrito (ética y derechos humanos) .....                     | 107 |
| Figura 3.23: Número de mensajes twitter según número de acceso al curso (ética y derechos humanos).....  | 108 |
| Figura 3.24: Número de intentos por rendir una evaluación según el tiempo utilizado (evaluación a distancia primer bimestre), (ética y derechos humanos) .....   | 109 |
| Figura 3.25: Número de intentos por rendir una evaluación según el tiempo utilizado (evaluación a distancia segundo bimestre), (ética y derechos humanos) .....  | 111 |
| Figura 3.26: Correspondencia del clúster 0 .....   | 128 |
| Figura 3.27: Correspondencia del clúster 1 .....   | 129 |

|  |     |
|--|-----|
| Figura 4.1: Modelo de datos del estudiante .....   | 142 |
| Figura A.1: Migración de DB .....  | 163 |
| Figura A.2: Migración de DB a Moodle .....   | 163 |
| Figura B.1: Consulta para ver todas las acciones que realiza el estudiante en el Eva ..... | 169 |
| Figura B.2: Consulta para ver registros de los cuestionarios .....                         | 170 |
| Figura C.1: Test de estilos de aprendizaje habilitado en cada curso .....                  | 176 |
| Figura D.1: Conexión a la base en WEKA con mysql .....                                     | 178 |
| Figura D.2: Parámetros de conexión .....   | 179 |
| Figura D.3: Conexión a la base MineríaAcciones con mysql .....                             | 179 |
| Figura D.4: Datos de los estilos de aprendizaje cargados en Weka .....                     | 180 |

## RESUMEN

La presente investigación está enfocada en la identificación de patrones de comportamiento relacionados con las acciones de los estudiantes que utilizan el entorno virtual de aprendizaje (EVA), se realizó un análisis previo de la base de datos correspondiente al periodo académico Abril2011/Agosto2001, para seleccionar exclusivamente las entidades que contienen información útil sobre las acciones que realizan los estudiantes. En la utilización de técnicas de minería se optó por la *clasificación: arboles de decisión, reglas de decisión y métodos bayesianos* éstas se utilizaron para clasificar los indicadores: *participación o interés en el curso y utilización de herramientas* por carrera siendo estas *informática y abogacía*, mientras que el *análisis de secuencias*, mediante la agrupación o agrupación en clústeres (clustering), fue utilizado para conocer los grupos de los estudiantes con características similares en relación al *indicador de la utilización de herramientas* y para encontrar el *estilo de aprendizaje dominante* y conocer cómo aprenden los estudiantes mediante el análisis de los estilos. Con los resultados obtenidos se detalló en un modelo de datos: *indicadores, atributos y estilos de aprendizaje del estudiante*.

# CAPÍTULO 1

## *Estado del arte*

## 1.1 Introducción

En el campo de Tecnologías para la Educación, la adaptación de sistemas educativos ofrece una forma avanzada en un ambiente de aprendizaje que intenta satisfacer las necesidades de los diferentes estudiantes, así como sistemas de construcción de un modelo de conocimiento del estudiante, las metas y preferencias del mismo. Aquí es donde juega un papel fundamental la aplicación de las técnicas de minería contribuyendo al descubrimiento de patrones de comportamiento de los estudiantes durante su interacción con el EVA de la UTPL.

Algunos entornos de aprendizaje virtual (EVA) son adecuados para la investigación del comportamiento de los alumnos en el aprendizaje. Uno de los más populares de éstos es Moodle ampliamente utilizado para la presentación de materiales de aprendizaje, así como para los debates entre los alumnos. Esta herramienta permite a un profesor no sólo informar materiales de aprendizaje de forma flexible, sino que también permite proporcionar la posibilidad de que los alumnos participen en las discusiones comunes, chats sincrónicos, crear sus blogs, archivos de revisión de vídeo de las conferencias, utilizar el correo electrónico, etc. Moodle es también una poderosa herramienta para el seguimiento de las acciones de los estudiantes y la interpretación de éstos resultados.

En esta primera parte de la tesis se desarrollará un estado del arte del aprendizaje, modelo de usuario, técnicas para el modelo de usuario, el proceso de construcción de modelos de estudiante, estilos de aprendizaje, proceso de minería de datos en la enseñanza a distancia, técnicas de minería de datos, entre otros.

Enfatizando en las actividades de aprendizaje individual, las cuales son las que realiza el estudiante directamente con el contenido definido por el profesor. No se requiere interacción con otros estudiantes.

Se analizará las mejores técnicas de construcción de modelos de interacción con el fin del descubrir el comportamiento de los estudiantes en base a las acciones que realizan en la plataforma Moodle.

Existe una estrecha relación entre los ámbitos de la Inteligencia Artificial (IA) y Aprendizaje Autónomo contando con las principales fuentes de minería de datos técnicas, procesos y métodos de la educación.

En (Baker, 2000), el autor establece las oportunidades de investigación en la IA y la educación en base a tres modelos de los procesos educativos: modelos como herramienta científica, se utilizan como medios para la comprensión y la predicción de algunos aspectos de una situación de la educación; modelos como componente: corresponde a alguna característica de la enseñanza o proceso de aprendizaje y se utiliza como un componente de un artefacto educativo, y modelos base para el diseño de objetos educativos: ayudar al diseño de herramientas informáticas para la educación, proporcionando metodologías de diseño y los componentes del sistema.

## **1.2 Modelo de usuario**

El Modelo de usuario, tiene como objetivo guardar los detalles de las preferencias y necesidades de usuario, dicha información es útil a la hora de adaptar el aprendizaje a éstos sistemas.

La arquitectura mundial propuesta por Benyon y De Bra indica que un sistema adaptativo debe tener tres partes esenciales (Martins, Faria, Vaz de Carvalho y Carrapatoso, 2008):

El modelo de usuario, que describe la información, el conocimiento, preferencias, etc., del usuario. Este componente permite extraer y expresar conclusiones sobre las características del usuario.

Que es lo que nos interesa para en base a las acciones que el usuario realice establecer las características que posee este, lo cual es el fundamento de la tesis.

**El modelo de dominio**, la función más importante de este modelo es proporcionar una estructura para la representación del conocimiento del dominio del usuario.

**El Modelo de Interacción**, representa y define la interacción entre el usuario y la aplicación, los datos que se almacenan en un modelo de interacción son utilizados para inferir en las características del usuario.

Esta definición describe la clasificación que se debe realizar de los perfiles de usuarios, en la cual se trataran atributos comunes, estos atributos en el caso concreto de los estudiantes de la modalidad Abierta de la UTPL son las acciones sobre objetos del entorno virtual de aprendizaje.

En (Gaudio, 2002), describe tres tipos de modelos:

- ❖ **Explícitos**, que se representan a través de reglas en base a un conocimiento declarado previamente.
- ❖ **Implícitos**, los cuales se basan en una serie de atributos que se obtienen con la ejecución de tareas de aprendizaje automático.
- ❖ **Híbridos**, que se basan en la combinación de las dos técnicas mencionadas anteriormente.

### 1.3 Técnicas para el modelo de usuario

Las técnicas de actualización de los modelos generalmente dependen del tipo de modelo del que se trate. Así por ejemplo, los modelos individualizados, los que se basan en el comportamiento de los usuarios, o los modelos a corto plazo, por lo general requieren actualizaciones dinámicas. Si el modelo contiene información a muy corto plazo entonces estamos hablando de un modelo de tareas, puesto que solo se refiere a la tarea que está

realizando el usuario en ese momento. En este caso, el modelo variara en cuanto el usuario cambie de tarea a realizar.

El tipo más básico de modelo que podemos considerar es el modelo estático con un usuario estándar (canónico). Este tipo de modelo se puede incorporar a un sistema fácilmente. Por el contrario, si lo que queremos es modelar cada uno de los usuarios individualmente entonces se necesitan actualizaciones dinámicas y son necesarios métodos explícitos que describan en qué manera el estado del modelo del usuario afectó a la respuesta que el sistema da al usuario.

Para terminar con esta clasificación encontramos dos fuentes de información diferentes para realizar el modelado de usuario. Puede recogerse directamente del usuario mediante la obtención de datos de formularios de entrada, o bien el sistema puede guardar las trazas de las interacciones del usuario y abstraer a partir de ellas la información que fuera necesaria (Horvitz, Breese, Heckerman, Hovel y Melse, 1998).

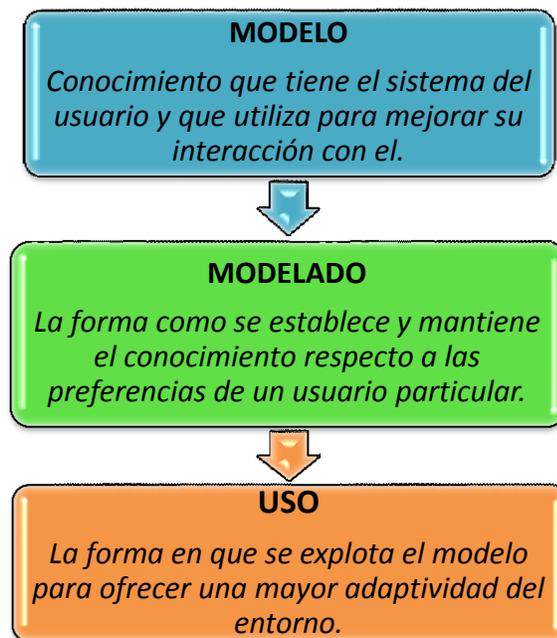


Figura 1.1: Esquema general de un modelo de usuario (Vélez, 2009)

## 1.4 Modelo de estudiante

En los sistemas de educación virtual el modelo de estudiante representa los gustos y necesidades del estudiante frente al proceso de aprendizaje y es utilizado ya sea para sugerirle nuevos cursos o para adaptar los contenidos y actividades de los cursos en los cuales está inscrito.

Estos modelos, generalmente son construidos a partir de información explícita (formas de registro, encuestas, cuestionarios) y/o información implícita (observando el comportamiento del estudiante). Holt define el modelo de estudiante como una representación de las creencias que tiene el sistema o aplicación acerca del usuario, y que por lo tanto, se considera una representación abstracta del estudiante en el sistema (Moreno, 2002).

Dicho modelo se utiliza para razonar acerca de las necesidades, preferencias o comportamiento futuro del usuario, así como para diagnosticar fallos en su interacción con el sistema.

En (Jiménez, Marcos y Fuentes, 2010), se encuentran, algunos métodos para representar el conocimiento del alumno, tales como:

- ❖ **Modelo overlay:** Es el método más utilizado, aquí el conocimiento del alumno se trata como un subconjunto del conocimiento del experto.
- ❖ **Modelo diferencial:** Divide el conocimiento del estudiante en dos categorías conocimiento que el estudiante debería conocer y conocimiento que se supone que el alumno no conoce.
- ❖ **Modelo de perturbación:** Se supone que el alumno posee conocimiento potencialmente diferente en cantidad y calidad respecto al de un experto.
- ❖ **Modelo de estado vs proceso:** Utilizan deducciones para generar predicciones.

## 1.5 Características a tener en cuenta en la construcción de un modelo del estudiante

En (Duque y Néstor, 2008), se mencionan que las características pueden ser propias al proceso educativo como son el nivel de comprensión de un tema, estilos de aprendizaje, gustos del usuario a la hora de aprender, características psicológicas como el estado de ánimo, las metas y por último su entorno como el clima y su estrato. Todas estas características relacionadas entre sí forman lo que se llama el modelo del estudiante, lo que permite construir un perfil para acomodar el sistema a las necesidades del usuario.

Para la construcción de un modelo del estudiante se deben tener en cuenta varios aspectos los cuales están especificados en la figura 1.2.

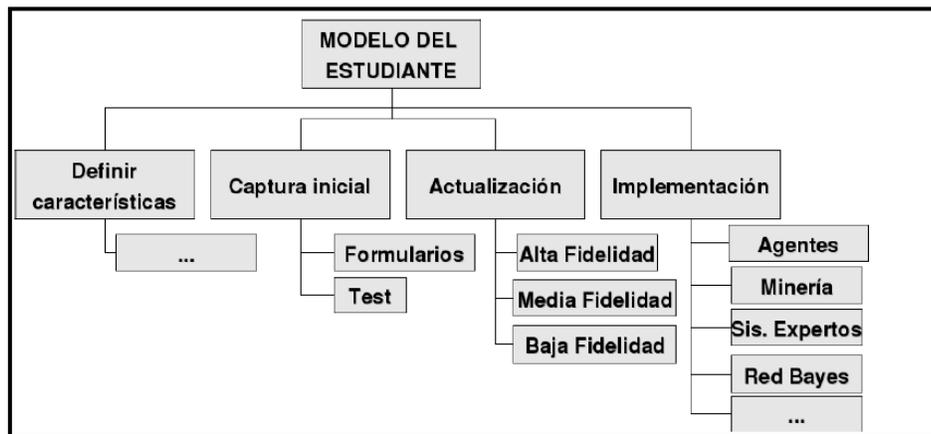


Figura 1.2: Aspectos a tener en cuenta en la construcción de un modelo del estudiante (Duque y Néstor, 2008)

En consecuencia el proceso de construcción del modelo seguirá la siguiente metodología:

- ❖ Definir las características del estudiante que afecten el proceso de enseñanza.
- ❖ Establecer las técnicas necesarias para realizar la captura inicial de los valores de cada una de las características definidas en el punto anterior.
- ❖ Seleccionar una forma de actualizar dinámicamente los valores de cada una de las características que lo ameriten.

- ❖ Seleccionar las técnicas más adecuadas para implementar el modelo en un sistema de educación virtual adaptativo.

Un modelo de Estudiante completo (Zaitseva y Boule, 2003), debe contener los siguientes aspectos:

- ❖ Nivel de aprendizaje.
- ❖ Características psicológicas.
- ❖ Estilo y velocidad de aprendizaje.
- ❖ Desarrollo de tarea
- ❖ Habilidad de aprendizaje
- ❖ Nivel de habilidades
- ❖ Métodos de las estrategias de enseñanza
- ❖ Gráfico de Conocimiento.

## 1.6 Técnicas para la construcción de modelo estudiante abierto

Partiremos de una serie de técnicas para el modelado de los estudiantes con el fin de mostrar como contribuyen para representar las creencias reales de los estudiantes, su imagen mental en relación con el conocimiento.

**Seguimiento de una Ruta (Model-Tracing):** (Anderson, Corbett, Koeindenger y Pelletier, 1995), la traza del conocimiento consiste en determinar qué sabe el alumno, incluyendo tanto el conocimiento correcto sobre el dominio como sus errores. La traza del modelo pretende analizar el procedimiento de resolución de problemas que utiliza el alumno. La traza del modelo resulta útil en sistemas que intentan dar respuesta a peticiones de ayuda del alumno y ofrecerle pistas e información cuando no sabe seguir resolviendo el problema. De hecho, para poder ayudar al alumno el sistema necesita ser capaz de analizar y criticar la solución en curso y tener una idea de que línea de razonamiento está siguiendo.

Por otro lado, la traza del conocimiento resulta útil para la evaluación del alumno y la toma de decisiones pedagógicas, como qué material/problema debe ser propuesto a continuación.

**Reconocimiento de un Plan (Plan Precognición):** (Anderson, Corbett, Koeindenger y Pelletier, 1995), en la técnica de reconocimiento de un plan se necesita cumplir dos condiciones: 1) que el conocimiento sea procedural y ordenado y 2) que todos o casi todos los estados físicos observables del estudiante puedan ser utilizados por el procedimiento de diagnóstico. En esta técnica el dominio se representa en un árbol, donde los nodos terminales son las acciones más primitivas, los nodos no-terminales incorporan los subobjetivos y la raíz el objetivo final.

**Seguimiento de asuntos (Issue-Tracing):** La técnica seguimiento de asuntos (Burton y Brown, 1982), está basada en el análisis de episodios cortos en la solución de un problema, dividiendo su observación en conjuntos de micro habilidades o asuntos que se emplean durante ese episodio. Este tipo de análisis no explica cómo estos asuntos interactúan, ni el papel que desempeñan en la solución global del problema, sólo interesa saber si se usan o no.

### 1.6.1 ¿Cómo construir un modelo del estudiante?

No hay un estándar acerca de que elementos deberían incluir un modelo del estudiante. Debería reflejar el estado actual del conocimiento del estudiante (Bretch y Jones, 1988), y eso significa, que no solamente debe contener los conceptos adquiridos sino también los errores conceptuales, el entendimiento parcial, etc. Un modelo del estudiante completo debería incluir todo el conocimiento relevante previo del estudiante, el progreso del estudiante, el estilo de aprendizaje preferido del estudiante, así como otros tipos de información relacionada a este. Implementar tal modelo sería una tarea formidable si no es que imposible; por lo tanto, un gran número de sistemas intentan modelar al estudiante solo en relación a la representación de la materia de interés (Holt, Dubs, Jones y Greer, 1994).

La construcción de un modelo del estudiante involucra delimitar: el “quién” está siendo modelado y cuál es la historia del estudiante, el “qué”, o las metas, planes, aptitudes, capacidades, conocimiento y creencias del estudiante; el “cómo” el modelo es adquirido y mantenido; y el “por qué”, incluyendo como serán las respuestas por parte del estudiante para darle asistencia, para proporcionarle retroalimentación, o para interpretar su comportamiento (Stauffer, 1996).

Las cuestiones anteriores son descritas a continuación (Finin, 1989).

### **¿Quién está siendo modelado?**

Hay dos cuestiones específicas que se necesitan considerar para evaluar quien está siendo modelado: el grado de especialización y el alcance temporal. El grado de especialización se refiere a que modelemos ya sea, individuos o clases de estudiantes. El alcance temporal hace referencia al alcance del conocimiento que mantendrá el estudiante a través del tiempo (historia del estudiante).

### **¿Qué está siendo modelado?**

El contenido del modelo del estudiante variará de aplicación a aplicación. Para construir un modelo se debe determinar que componentes son necesarios para mantener un modelo apropiado del estudiante. Los componentes representados pueden ser metas y planes del estudiante, capacidades, aptitudes y/o conocimiento o creencias.

### **¿Cómo puede ser adquirido y mantenido el modelo?**

En la adquisición uno está interesado acerca de las técnicas requeridas para aprender nuevos hechos acerca del estudiante. En el mantenimiento uno se preocupa por la habilidad de incorporar nueva información en el modelo del estudiante existente así como tratar con cualquier discrepancia.

### **¿Por qué está el modelo ahí?**

El modelo del estudiante puede ser utilizado de varias formas, sin embargo, el uso es a menudo dependiente de la aplicación. Por ejemplo, el modelo puede:

- ❖ Producir como respuesta información del estudiante.
- ❖ Dar al estudiante ayuda o consejos.
- ❖ Proporcionar retroalimentación al estudiante o
- ❖ Interpretar el comportamiento del estudiante.

Como todos los modelos, un modelo del estudiante tiene el fin de proveer información acerca del objeto modelado, en este caso, el estudiante que está utilizando un sistema de aprendizaje basado en computadora. El modelado del estudiante ocurre en tiempo de ejecución, cuando al sistema se le proporcionan datos acerca del estudiante; es entonces cuando el modelo del estudiante es creado (Self, 1992).

## **1.7 Estilos de aprendizaje y modelado de usuario**

Cada persona aprende de manera distinta a las demás: utiliza diferentes estrategias, aprende con diferentes velocidades y con mayor o menor eficacia incluso aunque tengan las mismas motivaciones, el mismo nivel de instrucción, la misma edad o estén estudiando el mismo tema (Woolfolk, 1996).

El concepto de estilo de aprendizaje es más amplio que el de estilos cognitivos puesto que además de comportamientos cognitivos, también incluye aspectos afectivos que indican las características y las maneras de percibir, interactuar y responder al contexto de aprendizaje por parte del estudiante.

Un modelo de estilo de aprendizaje clasifica a los estudiantes de acuerdo a donde ellos se ajustan en un número de escalas que consideran la forma como ellos reciben y procesan la información. (Felder y Brent , 2001).

Según (Felder y Silverman, 1988), el aprendizaje puede ser visto como una estructura educacional compuesta por dos pasos:

- ❖ **La recepción:** información externa (observable a través de sentidos) e información interna (adquirida introspectivamente) se convierte en válida para el estudiante, quien selecciona el material que procesará e ignora el resto.
- ❖ **El procesamiento de información:** simple memorización, razonamiento inductivo o deductivo, reflexión o acción e introspección o interacción con otros. El resultado es que el material, es aprendido o no es aprendido. De esta manera, dependiendo de la forma como se desarrolle el proceso enseñanza - aprendizaje, los dos pasos anteriormente descritos pueden facilitarse o dificultarse a un estudiante.

### **1.7.1 Adaptabilidad y preferencias de aprendizaje**

Existen diversos estilos de aprendizaje por lo que no hay una definición única, una manera de describirlo es: las distintas maneras en que un individuo puede aprender. Muchas de las teorías sobre estilos de aprendizaje (Felder y Brent, 2001), (Felder y Silverman, 1988), (Gardner, , 2003), (Kolb, 1984), (Herrmman, 1981), (Honney y Mumford, 2006), (Bandler y Grinder, 1975), (O'Connor y Seymour, 1995), se basan en la premisa que todas las personas emplean un método particular de interacción, aceptación y procesamiento de estímulos e información. De esta manera, en la enseñanza tradicional, las teorías basadas en la forma en que se procesan los estímulos y la información se resumen en la tabla 1.1

***Tabla 1.1: Clasificación de estilos de aprendizaje-enseñanza (Espinosa, Maza y López, 2009)***

| TEORIAS   | CLASIFICACION DE ESTILOS   |
|---|--|
| David Kolb  | <ul style="list-style-type: none"> <li>- Convergente</li> <li>- Divergente</li> <li>- Asimilador</li> <li>- Acomodador</li> </ul>  |
| Modelo de dominancia cerebral<br>Ned Herrmann   | Modo preferente de procesar información: <ul style="list-style-type: none"> <li>- Cortical izquierdo</li> <li>- Limbico izquierdo</li> <li>- Limbico derecho</li> <li>- Cortical derecho</li> </ul>                              |
| Sistemas representativos, Programación Neurolingüística (PNL).<br>Bandler, Grinder. O'Connor. | Sistemas representativos: <ul style="list-style-type: none"> <li>- Visual</li> <li>- Auditivo</li> <li>- Cenestésico</li> </ul>  |
| Estilos de aprendizaje (Felder y Silverman)   | <ul style="list-style-type: none"> <li>- Activo-reflexivo</li> <li>- Sensorial-intuitivo</li> <li>- Visual-verbal</li> <li>- Secuencial-global</li> </ul>  |
| Inteligencias múltiples de Gardner  | <ul style="list-style-type: none"> <li>- Lógico-matemático</li> <li>- Lingüístico-verbal</li> <li>- Corporal-kinestésico</li> <li>- Espacial</li> <li>- Interpersonal</li> <li>- Intrapersonal</li> <li>- Naturalista</li> </ul> |
| Método VARK   | <ul style="list-style-type: none"> <li>- Visual</li> <li>- Auditivo</li> <li>- Lector-Escritor</li> <li>- Kinestético</li> </ul>   |

Estas diferencias ejercen influencia en la manera que los alumnos interactúan con la información y determinan sus necesidades educativas. Por esta razón es conveniente tener en cuenta las diferentes características de los alumnos para adaptar los contenidos en función de estos y otros aspectos relevantes. Sin embargo el modelo de estilos de aprendizaje de Felder es uno de los modelos más ampliamente citados en trabajos referentes a la temática, para determinar el Estilo de Aprendizaje, propuso un cuestionario apropiado para estudios de ingeniería y de ciencias de la computación.

## **1.8 Ambientes de aprendizaje e interactividad**

Existen diversos conceptos acerca de ambientes de aprendizaje, lo más importante es que en ellos siempre se habla de las distintas posibilidades de comunicación y de los niveles de interactividad.

Desde la perspectiva de la educación en línea a través de ambientes de aprendizaje basados en materiales interactivos, un ambiente de aprendizaje es una entidad digital con características de diseño instruccional, que puede ser usada, reutilizada o referenciada durante el aprendizaje soportado en computadora, con el objetivo de generar conocimientos, habilidades y actitudes en función de las necesidades del alumno (Quiroz).

A partir de este punto de vista, la interacción constituye el núcleo de la actividad de un ambiente de aprendizaje, ya que el conocimiento es generado y construido. Construido en conjunto, exactamente porque se produce interactivamente entre dos o más personas que participan de él (Santarosa, 1999).

## **1.9 Características de los ambientes interactivos**

Un sistema de aprendizaje debe ser capaz de proporcionar al usuario facilidades como la navegación, retroalimentación y evaluación, así como la colaboración.

El aprendizaje a través de la colaboración requiere de la creación de un clima de confianza para que los estudiantes participen, expresen sus ideas y hagan sus aportaciones. Esto puede suceder solamente cuando se logra que los miembros del grupo tengan un compromiso con los demás.

(Muñoz y Muñoz), algunas estrategias que se deben aplicar en un ambiente de aprendizaje interactivo y por consecuencia se logre la colaboración, son:

- ❖ Fomentar la participación de los miembros.
- ❖ Generar preguntas enfocadas a los objetos de conocimiento.
- ❖ Aclarar y resolver dudas.
- ❖ Fomentar que los miembros construyan aportaciones para el grupo.
- ❖ Proporcionar una navegación en forma natural.
- ❖ Permitir mensajes instantáneos, charlas en línea.
- ❖ Registrar una bitácora de acciones por usuario.
- ❖ Controlar la cantidad de veces que participa un alumno.
- ❖ Registrar el rango de actividades en que participa un alumno.
- ❖ Contar con criterios y medios de evaluación y retroalimentación.

Otros aspectos que logran la interacción y generan aprendizaje, son:

- ❖ Crear escenarios para trabajos colaborativos, donde el profesor se encargue de facilitar, el proceso de enseñanza, estimulando a los alumnos a participar en los foros.
- ❖ Crear espacios lúdicos que permitan a los alumnos el fácil entendimiento de los materiales de estudio.
- ❖ Permitir al usuario navegar en el ambiente virtual y acceder a los materiales de estudio.
- ❖ Contar con páginas que se basen en modelos de realidad virtual, que contengan fotos, videos, animaciones.
- ❖ Prestar el servicio de discusión en línea, que sería similar al de las reuniones presénciales.

## **1.10 Datos de usuario**

Es la información de las características personales del usuario, datos que directamente se obtienen del sistema, de aquí la diferencia con los datos de uso que se relaciona con (interactiva) el comportamiento de un usuario, aquí ya estamos adentrándonos a la parte del

procesamiento de la información ya que es información generada de acuerdo al uso que se realice en la plataforma.

### **1.10.1 Datos demográficos**

Se refiere a los datos personales del usuario, entre éstos los que son de interés son los siguientes:

- ❖ Los datos de registro (nombre, dirección y número de teléfono)
- ❖ Los datos geográficos (ciudad, provincia y país)
- ❖ Características del usuario (edad, sexo, educación y estado civil)

### **1.11 Datos de uso**

Estos pueden ser directamente registrados, mediante el análisis de los datos que se observan. Dicha observación es técnicamente y varía entre éstos están los sistemas de hipermedia que se basan en HTML, y registran páginas que han sido solicitadas desde el servidor (incluye seguimiento de las rutas de navegación para los enlaces del sitio). Otros sistemas que tienen más control sobre la interacción, son los que utilizan applets de Java, puede guardar los datos de uso de acuerdo al nivel de clics del ratón y los movimientos. También están los sistemas que guardan logs<sup>1</sup> de la interacción del usuario con el mismo, así como el monitoreo de la interacción de los usuarios en un nivel de widgets<sup>2</sup>.

En el desarrollo de este modelo, la información de las interacciones con la plataforma por parte de los estudiantes es la que concretamente interesa, ya que notoriamente los datos de uso de foros, chats y videos van a permitir el nivel de actividad de éstos.

---

<sup>1</sup>Log. Un log es un registro oficial de eventos durante un periodo de tiempo en particular que generalmente se almacenan en archivos de texto o modelos relacionales diseñados para el caso.

<sup>2</sup>Widgets. Son pequeñas aplicaciones que se insertan en los sitios web brindando información o datos de importancia, como avisos sobre nuevos mensajes de correo electrónico.

A continuación se establece los datos de uso observables estos a su vez pueden ser utilizados para detallar las preferencias del usuario, hábitos y niveles de conocimientos.

### 1.11.1 Uso observables

En los sistemas adaptativos se puede identificar diferentes tipos de interacciones, ya que existen algunas maneras en las que los usuarios pueden interactuar con el sistema. Entre éstos tenemos los siguientes:

**Acciones Selectivas:** Se atribuye a (Kobsa, Koenemann y Pohl, 1999), el tipo más frecuente de interacción con los sistemas basados en la Web es hacer clic en un enlace. Al hacerlo, el usuario toma una decisión si los enlaces son competitivos de acuerdo a su perspectiva. Estas acciones selectivas pueden ser considerados como indicadores de varios tipos de datos de los usuarios:

- ❖ **Interés:** Con respecto a un objetivo de determinado usuario.
- ❖ **Desconocimiento:** De los usuarios con términos técnicos.
- ❖ **Preferencias:** Del usuario con respecto al tipo de objetos presentados.

**Ver el comportamiento temporal:** Determinación del tiempo de visualización de objetos hipermedia es más fácil si son escuchados (por ejemplo, de vídeo o audio) y si según una reacción de los usuarios (por ejemplo, solicitar, nueva página). El hecho de que el usuario esperó hasta el final de la presentación también puede considerarse como un indicador relativamente fuerte para el interés que manifiesta en determinado objeto.

**Clasificaciones:** Se refiere a valores que indican la relevancia o interés que el usuario da a objetos como (como documentos, artículos de noticias y productos). Normalmente, los usuarios realizan esto ya sea mediante una escala binaria (es decir, clasificar a un objeto como, por ejemplo, "interesante" o "no interesante") o una escala discreta. El último

tipo de escala de calificación es típicamente numérica (por ejemplo, todos los enteros de 0 a 5) o simbólico (con una asignación a una escala numérica).

**Otras acciones de confirmación y desaprobación:** Las acciones de confirmación son indicadores a menudo más fuertes que las acciones selectivas (Kobsa, Koenemann, y Pohl, 1999). Las acciones de confirmación, para los documentos como páginas Web, artículos de noticias o mensajes de correo electrónico, son interesantes para comprobar si el usuario hace un proceso como guardar un documento, imprimir un documento y marcar una página web aquí es donde se determinará si o no este tipo de acciones es observable por el sistema.

### 1.11.2 Regularidades de uso

Al observar las interacciones de los usuarios no llevan directamente a las adaptaciones por esto es necesario el tratamiento de los datos de uso para adquirir información sobre las preferencias de los usuarios, hábitos y niveles de experiencia sobre la cual puede ser el comportamiento del sistema adaptativo. En este apartado, se describen ejemplos típicos de uso de la información que se adquiere a partir de los datos de uso observados. Entre la distribución de tipos de datos realizado por (Kobsa, Koenemann, y Pohl, 1999) se encuentran los siguientes:

- **Frecuencia de uso.-** (Linton y Schaefer, 2000), compararon el uso individual a largo plazo de los comandos de Word con un grupo de datos de uso. Basados en esta comparación, los comandos que no son usados por usuarios individuales pero frecuentemente empleados por el grupo pueden ser puestos fuera del contexto de uso individual.

**Situación de la acción-correlaciones.-** Both, Mitchell y Maes (1995) describen un asistente de correo electrónico que generan recomendaciones de cómo el usuario podría tratar con los mensajes entrantes. Esas recomendaciones eran basadas en estadísticas

que expresan correlaciones generalizadas entre los correos anteriores (situaciones) y como el usuario los procesó (acciones).

**Secuencias de acción.-** Son analizadas para recomendar a la generación de macros de secuencias de acción de uso frecuente, predecir futuras acciones del usuario sobre la base de las acciones pasadas, y recomendar las acciones basadas en secuencias de acción frecuente de otros usuarios.

## 1.12 Tipo de interacción del usuario

Se pueden definir tres tipos de interacción, relacionadas al comportamiento o a la forma en que el estudiante interactúa con la plataforma:

***Interacción conformista,*** refleja un comportamiento, en que al estudiante le interesa el material que también le ha interesado a la mayoría de estudiantes. Esto quiere decir que su nivel de interacción con las herramientas de la plataforma es *BAJO* y se debería presentar recomendaciones básicas, con el fin de incentivar al estudiante a utilizar con mayor frecuencia las herramientas de la plataforma y así incrementar su nivel de participación y su experiencia como usuario.

***Interacción consciente,*** en este tipo de interacción se puede observar una mayor utilización de las herramientas, por lo que se concluye que el usuario tiene bien definidos cuáles son sus intereses en cuanto al material, que quiere consultar. Esto representa a un estudiante que le interesa revisar los foros y los anuncios que el profesor público, por lo que el tipo de recomendación estaría fundamentada en un nivel más avanzado y que se incite al estudiante a participar directamente con las herramientas de la plataforma. Dicha interacción está relacionada con un nivel de interacción *MEDIO*.

**Interacción autónoma**, la interacción autónoma se relaciona con un nivel de interacción *ALTO*, ya que los intereses del usuario están basados en la consulta de material novedoso, participación frecuente en los foros y sobretodo experiencia en la utilización de las herramientas (Alberca, Loja, Valdiviezo, Jimenéz, y Agila, 2008).

### 1.13 Minería de datos para el modelo del estudiante

El objetivo en todo proceso de Minería de Datos es obtener patrones de interés para el usuario final. Para lograrlo, es necesario preparar correctamente a los datos para procesarlos, elegir un método adecuado para extraer los patrones deseados y finalmente, determinar cómo evaluar los patrones encontrados (Olmos y Gonzáles, 2007).

**Especificación del problema:** En esta etapa se determinan las fuentes de datos y el tipo de información a utilizar. Es la etapa donde los datos relevantes para el análisis son extraídos desde la o las fuentes de datos.

**Preprocesamiento:** Esta etapa consiste en la preparación y limpieza de los datos extraídos desde las distintas fuentes de datos en una forma manejable, necesaria para las fases posteriores. En esta etapa se utilizan diversas estrategias para manejar datos faltantes o en blanco, datos inconsistentes o que están fuera de rango, obteniéndose al final una estructura de datos adecuada para su posterior transformación.

**Transformación:** Consiste en el tratamiento preliminar de los datos, transformación y generación de nuevas variables a partir de las ya existentes con una estructura de datos apropiada. Aquí se realizan operaciones de agregación o normalización, consolidando los datos de una forma necesaria para la fase siguiente.

**Data Mining:** Es la fase de modelamiento propiamente tal, en donde métodos inteligentes son aplicados con el objetivo de extraer patrones previamente

desconocidos, válidos, nuevos, potencialmente útiles y comprensibles y que están contenidos u “ocultos” en los datos.

**Interpretación y Evaluación:** Se identifican los patrones obtenidos y que son realmente interesantes, basándose en algunas medidas y se realiza una evaluación de los resultados obtenidos ( WebMining Consultores, 2011).

En la figura 1.3, se aprecia brevemente las fases de cómo del proceso de descubrimiento de conocimiento en base de datos (KDD, por sus siglas en inglés)

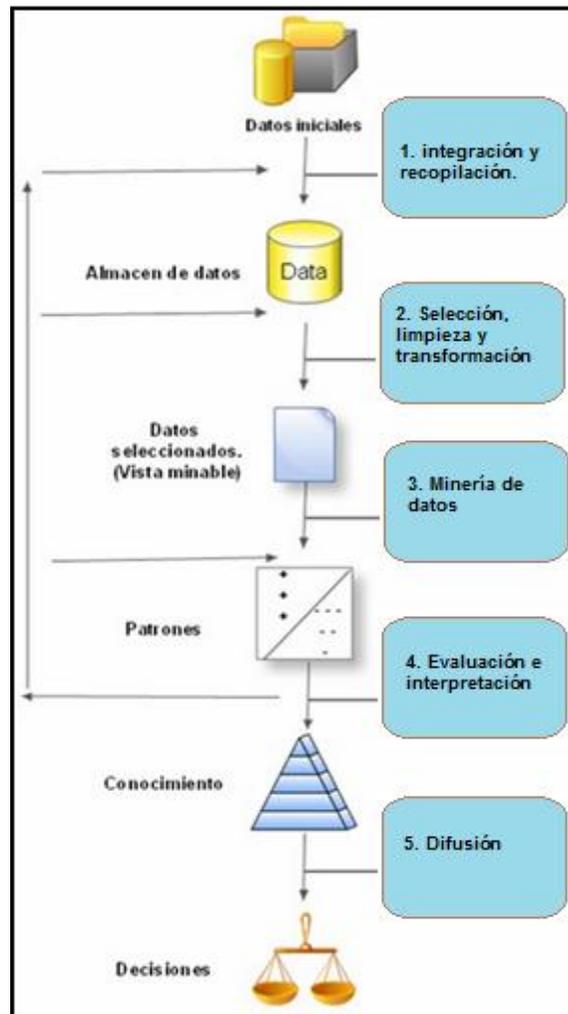


Figura 1.3: Fases del KDD (Hernández, Ramírez y Ferri, 2008)

## 1.14 El proceso de minería de datos en la enseñanza a distancia

Datos de los métodos de aplicación de minería en la educación a distancia consiste en cuatro pasos figura 1.4:

1. La acumulación de datos.
2. Preparación de los datos acumulados.
3. Aplicación de métodos de minería de datos para los datos seleccionados.
4. Interpretación y análisis de los datos seleccionados (Romero, Ventura y García, 2007).

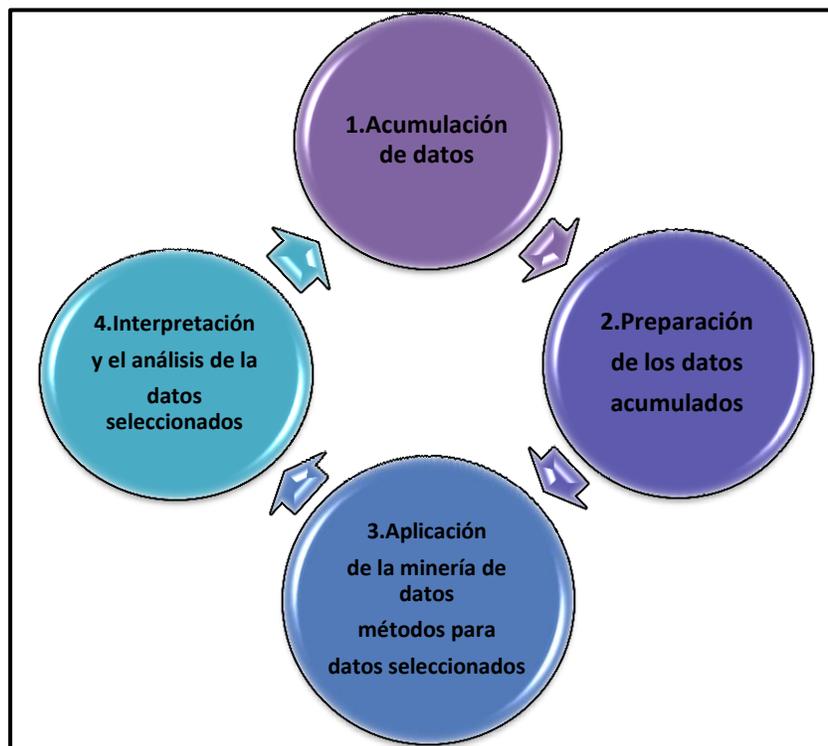


Figura 1.4: Pasos de los métodos de aplicación de minería de datos en la enseñanza a distancia (Romero, Ventura, y García, 2007)

## 1.15 Técnicas de minería de datos

Las técnicas de minería de datos se aplican a un conjunto de datos para obtener unos resultados entre las técnicas de Minería de datos utilizadas están las siguientes (Bressán, 2003):

- ❖ **Análisis Preliminar de datos usando Query tools:** Es el primer paso de un proyecto de MD, se aplica una consulta SQL<sup>3</sup> al conjunto de datos, para rescatar algunos aspectos visibles antes de aplicar las técnicas.
- ❖ **Técnicas de Visualización:** Son aptas para ubicar patrones en un conjunto de datos, puede usarse al comienzo de un proceso de MD para determinar la calidad de los datos.
- ❖ **Redes neuronales artificiales:** Son modelos predecibles, no lineales que aprenden a través del entrenamiento.
- ❖ **Reglas de Asociación:** Establecen asociaciones en base a los perfiles de los clientes sobre los cuales se realiza la MD.
- ❖ **Algoritmos Genéticos:** Son técnicas de optimización que usan procesos tales como combinaciones genéticas y mutaciones, etc.
- ❖ **Métodos Bayesianos:** Ofrecen un análisis cualitativo de los atributos y valores que pueden intervenir en el problema, además de que manifiesta también importancia cuantitativa de esos atributos. En el aspecto *cualitativo* se puede representar cómo se relacionan esos atributos ya sea en una forma causal, o señalando simplemente de la correlación que existe entre esas variables (o atributos), mientras que *cuantitativamente*, da una medida probabilística de la importancia de esas variables

---

<sup>3</sup>SQL, lenguaje de consulta estructurado.

en el problema por lo tanto una probabilidad explícita las hipótesis que se formulan (Mitchell y Tom, 1997).

❖ **Algoritmo BayesNet**

- Representan el conocimiento cualitativo del modelo mediante un grafo dirigido acíclico. Este conocimiento se emite en la definición de relaciones de independencia/dependencia entre las variables que componen el modelo, mismas relaciones que comprenden desde una dependencia completa hasta una dependencia funcional entre las variables.
- Expresan en forma numérica las relaciones entre variables, esta parte suele especificarse mediante distribuciones de probabilidad como una medida de la creencia que se tiene sobre las relaciones entre las variables de modelo (Hernández, Ramírez y Ferri, 2008).

❖ **Algoritmo NaiveBayes**

- Asume que todos los atributos son condicionalmente independientes cuando la clase es conocida. Esto implica que la estructura de la red es la más simple, ya que los únicos arcos en el grafo dirigido son los que relacionan la variable clase con cada una de las características. La ventaja es que tiene que aprender un menor número de parámetros que existen en las relaciones de independencia, ya que son las más simples y esto condiciona una menor tasa de error (Duda, Hart y Stork, 2001).

- ❖ **Árbol de Decisión:** Son estructuras que representan conjuntos de decisiones, y estas decisiones generan reglas para la clasificación de un conjunto de datos. (Unne Faccena, 2003).

#### ❖ **Algoritmo J48**

- Permite trabajar con valores continuos para los atributos, separando los posibles resultados en dos ramas.
- Genera un árbol de decisión a partir de los datos mediante particiones realizadas recursivamente, según la estrategia de profundidad-primero.
- Escoger un rango de medida apropiado.
- Manejo de datos de entrenamiento con valores faltantes.

#### ❖ **Algoritmo REPTree**

- Es un método de aprendizaje rápido mediante árboles de decisión. Construye un árbol de decisión usando la información de varianza y lo poda usando como criterio la reducción del error. Clasifica valores para atributos numéricos una única vez. Los valores que faltan se obtienen seccionando las correspondientes instancias (Duque y Orozco, 2011).

❖ **Reglas de decisión:** Entre los algoritmos que conforman este tipo de metodología encontramos el **JRip** este genera un listado de reglas las cuales son obtenidas a partir de listas de decisión que se evalúan en orden para encontrar la regla más eficiente para la información que se está manipulando.

❖ **Algoritmos particionales:** Estos algoritmos asumen un conocimiento a priori del número de clúster en que debe ser dividido el conjunto de datos, llegan a una división en clases que optimiza un criterio predefinido o función objetivo. Entre los algoritmos que emplean esta técnica podemos mencionar:

#### ❖ Algoritmo K-Means

- Es uno de los más simples y conocidos algoritmos de agrupamiento, sigue una forma fácil y simple para dividir una base de datos dada en k grupos (fijados a priori).
- La idea principal es definir k centroides (uno para cada grupo) y luego tomar cada punto de la base de datos y situarlo en la clase de su centroide más cercano. El próximo paso es recalcular el centroide de cada grupo y volver a distribuir todos los objetos según el centroide más cercano. El proceso se repite hasta que ya no hay cambio en los grupos de un paso al siguiente (Duda, Hart y Stork, 2001).
- El problema del empleo de éstos esquemas es que fallan cuando los puntos de un grupo están muy cerca del centroide de otro grupo ver ejemplo en (Karypis, Han, y Kumar, 1999), también cuando los grupos tienen diferentes tamaños y formas.

#### ❖ Algoritmo FarthestFirst

- La función principal de este algoritmo es seleccionar aleatoriamente una instancia que pasa a ser el centro del clúster. Se calcula la distancia entre cada una de las instancias y el centro. La distancia que se encuentre más alejada del centro más cercano es seleccionada como el nuevo centro del clúster. Este proceso se repite hasta alcanzar el número de clúster buscado (Pinar, 2007).

Para efectos del trabajo de esta tesis, se utilizó como punto de partida la técnica de análisis preliminar de datos usando *Query tools*, útil a la hora de saber cuál es la distribución de los

valores posibles de los atributos que interesan sobre las acciones que realizan los estudiantes sobre la plataforma.

## 1.16 Construcción del modelo de predicción

Es en la construcción del modelo donde vemos mejor el carácter interactivo, del proceso de KDD, ya que será necesario explorar modelos alternativos hasta encontrar aquel que resulte más útil para resolver nuestro problema. El proceso de construcción de modelos predictivos requiere tener bien definidas las etapas de entrenamiento y validación para asegurar que las predicciones serán robustas precisas (Hernández, Ramírez y Ferri, 2008).

La Teoría de la Actividad es un método para estudiar el comportamiento humano basado en la actividad como la unidad de análisis (Leont'ev, 1978). La actividad se conforma por un sujeto, un objeto, acciones y operaciones. El sujeto es el responsable de la actividad. El objeto representa la meta que la actividad se propone satisfacer. Las acciones son las partes en que se descompone la actividad, mientras que las operaciones son tareas que satisfacen condiciones.



Figura 1.5 Relación adquisición de conocimiento – fenómeno causal – actividad (Ayala, A., 2008)

La Teoría de la Actividad establece cinco principios: objeto, mediación, jerarquía, desarrollo y el ciclo de exteriorización – interiorización. En la Teoría de la Actividad se postula que la actividad humana está guiada por la anticipación, la cual es el motivo de la actividad. La anticipación de futuros eventos es el principio fundamental de la reflexión anticipada. Una vez que se realiza la actividad, surge un flujo de retroalimentación que motiva comparar el

resultado contra la predicción. Como consecuencia de la evaluación, cualquier imprecisión detectada produce una oportunidad de aprendizaje que enriquece la experiencia del individuo.

Dentro del marco conceptual de la Teoría de la Actividad, la propuesta analiza el proceso de enseñanza-aprendizaje como un fenómeno causa-efecto, que se corresponde con el principio de exteriorización-interiorización, tal como se ilustra en la figura 1.5. El modelo del estudiante recrea un modelo mental del proceso de enseñanza-aprendizaje teniendo claro que la educación centrada en el estudiante procura que el individuo adquiera conocimientos sobre un dominio específico, el objeto, mediante la reflexión anticipada, la cual es inferida por el mapa cognitivo que representa al modelo del estudiante.

## **1.17 Trabajos relacionados**

Este punto tiene como objetivo, encuadrar la tesis dentro de investigaciones sobre la minería de datos en un entorno virtual de aprendizaje. A continuación se describirán los diferentes enfoques sobre las diferentes proximidades al modelado de las acciones de los estudiantes sobre un EVA.

### **1.17.1 Inferencia en el modelo de usuario**

El modelo de usuario propuesto toma los datos de los estilos de aprendizaje de los alumnos a través del cuestionario ILS<sup>4</sup>. Las propiedades del curso se adaptan a esas características iniciales que pueden ser modificadas de forma dinámica en función de las acciones del usuario. Conforme la cantidad de información (acciones del usuario) va creciendo, el modelo de usuario puede ir cambiando o no (Paredes y Rodríguez, 2004).

Las acciones y los resultados a tener en cuenta en la interacción del estudiante con el sistema son:

---

<sup>4</sup>ILS. Índice de estilos de aprendizaje.

- ❖ Las tareas que han sido realizadas y en qué orden.
- ❖ El tiempo que el estudiante ha tardado en realizar las tareas.
- ❖ El porcentaje de ejercicios resueltos correctamente.

**Características individuales**, éstos modelos representan características particulares de cada usuario. Durante el trabajo del usuario con el sistema estas características pueden cambiar, entonces el objetivo de éstos modelos es seguir las acciones del usuario y mantener actualizado dicho modelo. De acuerdo a (Brusilovsky y Millan, 2008) Las cinco características más populares son: nivel de conocimiento, interés, metas experiencias previas y rasgos particulares.

### **1.17.2 La minería de datos en los sistemas de gestión de cursos: Caso de estudio de Moodle tutorial**

La aplicación de minería de datos en los sistemas de e-learning es un ciclo repetitivo en el que el conocimiento extraído debe entrar en el bucle del sistema y la guía, facilitando y mejorando el aprendizaje en su conjunto, no sólo convertir los datos en conocimiento, sino también filtrar el conocimiento extraído de la toma de decisiones (Romero, Ventura, y García, 2007). El proceso de e-learning de minería de datos se compone de los mismos cuatro pasos en el proceso de minería de datos en general de la siguiente manera:

**Recopilar datos:** El sistema LMS es utilizada por los estudiantes y el uso de la información y la interacción se almacena en la base de datos. En este artículo vamos a utilizar los datos de los estudiantes el uso del sistema Moodle.

**Preprocesar los datos:** Los datos se limpian y se transforman en un formato adecuado para ser explotado. Con el fin de preprocesar los datos de Moodle, podemos usar una herramienta de administración de bases de datos o alguna herramienta de preprocesamiento específicas.

**Aplicar la minería de datos:** Los algoritmos de minería de datos se aplican para crear y ejecutar un modelo que se descubre y resume los conocimientos de interés para el usuario (profesor, estudiante, administrador, etc.). Con el fin de hacerlo podemos utilizar una herramienta de minería de datos específica, comercial o libre.

**Interpretar, evaluar y desplegar los resultados:** Los resultados obtenidos son modelos interpretados y utilizados por el profesor para tomar nuevas medidas, puede utilizar la información descubierta para tomar decisiones sobre los alumnos y las actividades del curso de Moodle con el fin de mejorar el aprendizaje de los estudiantes.

### **1.17.3 Minería de datos al estudiante para caracterizar el comportamiento similar en grupos de espacios de colaboración estructurada**

(Talavera y Gaudioso, 2004), proponen realizar clústeres para descubrir patrones que reflejen el comportamiento de los usuarios en espacios de colaboración. Para esto realizaron un análisis donde la característica principal utilizada para formar los grupos fueron las respuestas a los intereses y habilidades. Además en los experimentos que ellos realizan siguen un esquema con una característica externa que indica si un alumno ha aprobado o reprobado el curso, lo hacen con el fin de determinar manualmente el número más adecuado de los clúster.

Entre la solución que ellos utilizan para resolver el problema es el algoritmo EM, el cual toma como entrada los datos y se necesita el número de grupos, los resultados, los parámetros del modelo y las probabilidades posteriores para cada instancia, la probabilidad de que cada instancia se ha generado por el grupo. Teniendo en cuenta que el algoritmo EM asume que el número de de grupos se conoce de antemano, por lo que no se aborda directamente el problema de encontrar la estructura del modelo (número de clúster).

#### **1.17.4 Aplicación de minería de datos con una herramienta de software libre en la evaluación del rendimiento académico de los alumnos de la carrera de Sistemas de la FACENA-UNNE**

(Dapozo, Porcel, Lopez, Bogado y Bargiela), realizan un trabajo en el que hacen uso de técnicas de minería de datos que les permitió, a través de un clasificador, determinar el rendimiento académico de los alumnos de la carrera de Licenciatura en Sistemas de Información, para cumplir con esto llevaron a cabo un estudio comparando los diferentes algoritmos clasificadores disponibles en WEKA, analizando los resultados que se obtuvieron como resultado de la aplicación de cada uno de ellos, los datos que ellos utilizaron en la fase del análisis fue a partir de la base que alojaba toda la información sistematizada de los alumnos de la Facultad de Ciencias Exactas de la UNNE tomando particularmente en cuenta los siguientes datos:

- ❖ La cantidad de exámenes finales rendidos (número de intentos)
- ❖ La cantidad de exámenes finales aprobados
- ❖ La cantidad de exámenes finales reprobados

En base a esos valores, ellos generaron tres categorías que sirvieron para identificar a los alumnos que:

- ❖ En el año de ingreso no rindieron ninguna materia
- ❖ Los que si rindieron pero no aprobaron ninguna
- ❖ Los que rindieron y aprobaron por lo menos una materia

Estas categorías, según los intentos realizados y resultados obtenidos en los exámenes finales de las materias del primer año de la carrera les permitieron estimar el rendimiento académico de los alumnos.

## CAPÍTULO 2

### *Análisis del EVA de la UTPL*

## 2.1 Introducción

Un Sistema basado en web constituye una herramienta de apoyo para instituciones educativas que desean incorporar esquemas de trabajo a distancia o abierta, proporcionando espacios de comunicación, colaboración y contenidos para que profesor-estudiante interactúen virtualmente construyendo en conjunto, conocimientos sobre temáticas establecidas a través de esquemas y aproximaciones al aprendizaje social. Un entorno virtual de aprendizaje (EVA) es una herramienta desarrollada sobre Moodle que se encuentra implementado en la Universidad Técnica Particular de Loja (UTPL), precisamente para un sistema de educación a distancia, que permite a los estudiantes la interacción en este entorno desarrollando su proceso de aprendizaje, facilitando la personalización en sus tiempos de estudio así como la posibilidad de estar en permanente comunicación de manera privada con el profesor y acceder a las diferentes fuentes de información que éste les facilite sobre la asignatura, algunas de estas son: anuncios, ejercicios, exámenes, cuestionarios, consultas y para realizar tutorías virtuales usando los foros, el chat y correo electrónico.

## 2.2 Moodle

MOODLE (Entorno Modular de Aprendizaje Dinámico Orientado a Objetos), es una plataforma que se instala en un servidor Web de Internet y que permite recoger toda la información y documentación necesaria para poder impartir un curso a distancia. Una vez instalado en un servidor se pueden crear diversidad de cursos los cuales pueden agruparse en categorías. Cada uno de éstos cursos pueden ser tutorizados por uno o varios profesores, los cuales pueden ir haciendo el seguimiento de sus alumnos a distancia. Resulta esencial para crear “objetos de aprendizaje” o “unidades didácticas” y para fomentar el auto aprendizaje y el aprendizaje cooperativo. También es la herramienta ideal para gestionar la organización de las comunidades educativas y permitir la comunicación y el trabajo en red entre sus distintos integrantes y con otros centros (Ros, 2008).

### 2.2.1 Herramientas de apoyo para el aprendizaje

Las herramientas que se detallarán a continuación pueden ser usadas para tener un trato más directo con los alumnos así como ayudar a los profesores a desarrollar en ellos otras habilidades (González de Felipe).

#### 1 COMUNICACIÓN

-  Chat
-  Consulta
-  Foro
-  Encuesta

#### 2 CONTENIDO

-  Glosario
-  Lección
-  Etiqueta
-  Recurso
-  Wiki
-  Scorm

### 3 EVALUACIÓN

 Cuestionarios

 Tarea

 Taller

## 2.2.2 Herramientas de apoyo para el aprendizaje en el EVA de la UTPL

La implementación del EVA en la UTPL, ofrece a los estudiantes en formación la oportunidad de reforzar aprendizaje brindando medios de comunicación, contenido y evaluación. Las cuales se concretan en herramientas específicas que son utilizadas para posibilitar la interacción estudiante-estudiante, estudiante-profesor y estudiante-universidad, las que se encuentran alojadas para su uso son las siguientes:

### COMUNICACIÓN

✓ Foro/Grupo

Actividad para el debate entre usuarios de un curso.

|                        |   |
|------------------------|---|
| <b>Características</b> | <ul style="list-style-type: none"> <li>✓ Hay diferentes tipos de foros disponibles: exclusivos para los profesores, de noticias del curso y abiertos a todos.</li> <li>✓ <b>Foros de debate sencillo:</b> Simplemente un intercambio de ideas sobre un solo tema, todo en una página, respondiendo a un único planteamiento inicial. Útil para debates cortos y muy concretos.</li> <li>✓ <b>Foro Normal, para uso general:</b> Foro abierto donde cualquiera puede empezar un nuevo tema de debate cuando quiera. Este es el foro más adecuado para uso general.</li> <li>✓ <b>Foro Un Debate por Persona:</b> Cada persona puede plantear un nuevo tema de debate (y todos pueden responder). Esta modalidad es útil cuando quieres que cada alumno o miembro del grupo empiece una discusión sobre, digamos, sus conclusiones sobre el tema de la semana,</li> </ul> |
|------------------------|---|

|           |   |
|-----------|---|
|           | y que todos los demás le respondan.   |
| ✓ Twitter | Es utilizado por el estudiante para manifestarle al profesor o a sus compañeros alguna inquietud que tuviesen con respecto a la asignatura de la que sean partícipes. |

## CONTENIDO

|                        |  |
|------------------------|--|
| ✓ Glosario             | Permite crear y mantener una lista de definiciones, como un diccionario.   |
| <b>Características</b> | ✓ Muestra en su interior la definición o descripción de cualquiera de los glosarios que se haya definido en el curso.  |
| ✓ Etiqueta             | Sirve para añadir una descripción mediante texto e imágenes en la página principal de un curso.  |
| ✓ Recurso              |  |
| <b>Características</b> | ✓ Dentro de los recursos de apoyo a la formación se aprovecha las ventajas que brinda el sitio de youtube y slideshare, como la de permitir embeber los videos y presentaciones en blogs y portales, pudiendo así integrar los videos y presentaciones para las tutorías académicas tanto del primer como del segundo bimestre en las asignaturas disponibles en el EVA, por cada periodo académico. |

## EVALUACIÓN

|         |  |
|---------|--|
| ✓ Tarea | Actividad de trabajo en grupo que permite la evaluación entre estudiantes. |
|---------|--|

|   |  |
|---|--|
| <p><b>Características</b></p>   | <ul style="list-style-type: none"> <li>✓ Permite la evaluación de documentos entre iguales, y el profesor puede gestionar y calificar la evaluación.</li> <li>✓ Admite un amplio rango de escalas de calificación posibles.</li> <li>✓ El profesor puede suministrar documentos de ejemplo a los estudiantes para practicar la evaluación.</li> <li>✓ Es muy flexible y tiene muchas opciones.</li> </ul>  |
| <p>✓ Cuestionario</p> <p>Permite la realización de exámenes de diferente tipo: respuesta múltiple, verdadero/falso y respuestas cortas.</p> |  |
| <p><b>Características</b></p>   | <ul style="list-style-type: none"> <li>✓ Los profesores pueden definir una base de datos de preguntas que podrá ser reutilizadas en diferentes cuestionarios.</li> <li>✓ Las preguntas pueden ser almacenadas en categorías de fácil acceso, estas categorías pueden ser "publicadas" para hacerlas accesibles desde cualquier curso del sitio.</li> <li>✓ Los cuestionarios se califican automáticamente, y pueden ser recalificados si se modifican las preguntas.</li> <li>✓ Los cuestionarios pueden tener un límite de tiempo a partir del cual no estarán disponibles.</li> <li>✓ El profesor puede determinar si los cuestionarios pueden ser resueltos varias veces y si se mostrarán o no las respuestas correctas y los comentarios.</li> <li>✓ Las preguntas y las respuestas de los cuestionarios pueden ser mezcladas (aleatoriamente) para disminuir las copias entre los alumnos.</li> <li>✓ Las preguntas pueden crearse en HTML y con imágenes.</li> <li>✓ Las preguntas pueden importarse desde archivos de texto externos.</li> </ul> |

## 2.3 Acciones de los estudiantes en el eva

A continuación en la siguiente tabla 2.1 se describe las principales acciones que el estudiante realiza sobre la plataforma.

*Tabla 2.1: Atributos de las acciones del usuario sobre la plataforma*

| <b>module</b> | <b>action</b>   | <b>DESCRIPCIÓN</b>  |
|---------------|-----------------|---|
| COURSE        | VIEW            | Número de accesos al curso seleccionado   |
| ASSIGNMENT    | VIEW            | Número de veces que el usuario accede a las tareas enviadas por el profesor.  |
| FORUM         | ADD_POST        | Número de veces que el usuario envía un POST <sup>5</sup> , a un foro es decir un mensaje que da respuesta a alguna pregunta que realizo el profesor. |
| FORUM         | VIEW_DISCUSSION | Número de veces que el usuario revisa las discusiones dentro de un foro.  |
| FORUM         | VIEW_FORUM      | Número de veces que el usuario revisa o accede a los foros de discusión.  |
| FORUM         | UPDATE_POST     | Número de veces que el usuario actualiza el POST que se envió al foro.  |
| FORUM         | ADD_DISCUSSION  | Número de veces que el usuario agrega un tema de discusión o debate.  |
| RESOURCE      | VIEW            | Número de veces que el usuario accede descarga recursos <sup>6</sup> subidos por el profesor.   |
| USER          | UPDATE          | Número de veces que el usuario actualiza los datos de su perfil.  |
| USER          | VIEW            | Número de veces que el usuario revisa o accede a su perfil.   |
| UPLOAD        | UPLOAD          | Número de veces que el usuario sube una tarea.  |
| MESSAGE       | WRITE           | Número de mensajes que el usuario ha escrito.   |

<sup>5</sup> POST se refiere a la acción de enviar un mensaje hacia el servidor principal.

|         |                       |   |
|---------|-----------------------|---|
| MESSAGE | HISTORY               | Número de visitas al historial de los mensajes.                     |
| QUIZ    | VIEW                  | Número de veces que el usuario revisa o accede a los cuestionarios. |
| QUIZ    | TIMESTART, TIMEFINISH | Tiempo que tarda en la resolución de un cuestionario.               |
| QUIZ    | ATTEMPT               | Número de veces que el usuario intenta resolver un cuestionario.    |

## 2.4 Modelo entidad/relación del eva

Luego de un previo análisis de la base de datos del EVA. En la figura 2.1 se ha seleccionado concretamente las entidades que contienen información útil sobre las acciones que los estudiantes realizan sobre el EVA de la UTPL.

- ❖ mdl\_enrol\_utpl
- ❖ mdl\_user
- ❖ mdl\_log
- ❖ mdl\_twt\_utpl
- ❖ mdl\_quiz\_attempt

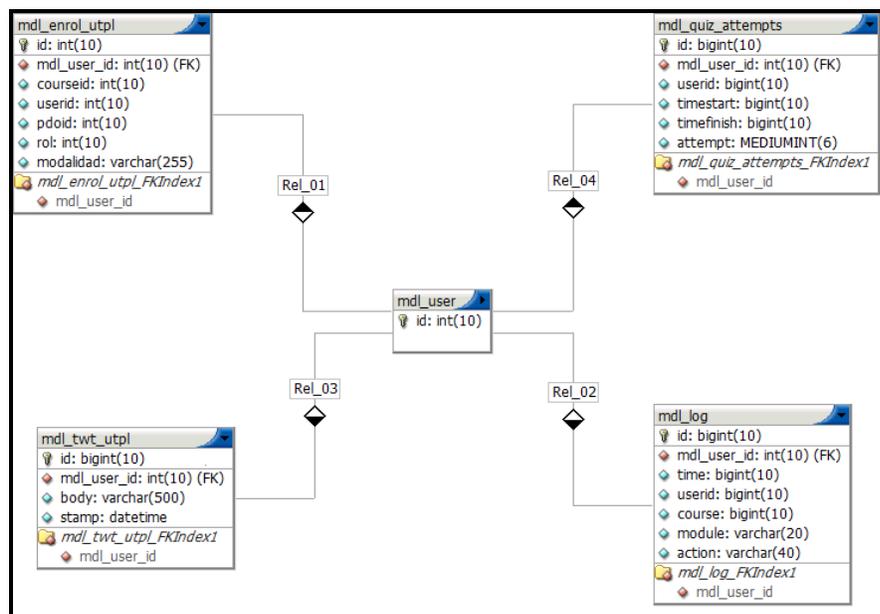


Figura 2.1: E/R que contienen información de las acciones de los estudiantes.

## 2.4.1 Entidades de moodle relacionadas con las acciones de los estudiantes en el eva

Las entidades que se utilizaron para conocer la información del modelo de estudiante respecto a las acciones que éste realiza sobre el EVA se detallaran a continuación:

*Tabla 2.2: mdl\_enrol\_utpl*

| a) mdl_enrol_utpl | DESCRIPCIÓN  |
|-------------------|--|
| Atributo          |  |
| ✓ id              | Identifica enrol_utpl  |
| ✓ courseid        | Identificador del curso  |
| ✓ userid          | Identificador del usuario  |
| ✓ pdoid           | Identificador del periodo académico  |
| ✓ rol             | Identifica si el usuario es docente o estudiante   |
| ✓ modalidad       | Indica si los estudios que cursa el estudiante son en modalidad 'Clásica' o 'Distancia'. |

La tabla 2.2. Es donde se recogen todos los atributos identificadores únicos de las entidades de mayor relevancia facilitándonos el relacionarlas entre ellas, para nuestra labor la extracción de información útil, estas entidades son: mdl\_course, mdl\_user, mdl\_periodo. Además de contar con el atributo que determina el rol del usuario y modalidad de estudios.

*Tabla 2.3: mdl\_user*

| b) mdl_user | DESCRIPCIÓN               |
|-------------|---------------------------|
| Atributo    |                           |
| ✓ id        | Identificador del usuario |

La tabla 2.3. Contiene información privada del estudiante es por esto que solo se tomo el atributo del identificador, ya que el resto información es irrelevante.

Tabla 2.4: mdl\_log

| c) mdl_log   | DESCRIPCIÓN  |
|--|--|
| <b>Atributo</b>  |  |
| <ul style="list-style-type: none"> <li>✓ id</li> <li>✓ time</li> <li>✓ userid</li> <li>✓ course</li> <li>✓ module</li> <li>✓ action</li> </ul> | <b>Acción sobre un objeto:</b> Que acción se realizó sobre un objeto.  |
| <b>EJEMPLOS</b>  |  |
| <ul style="list-style-type: none"> <li>✓ userid</li> <li>✓ course</li> <li>✓ module course</li> <li>✓ action-view</li> </ul>                   | <b>Número de accesos a un curso:</b> Las veces que un usuario accede a la vista de un curso.   |
| <ul style="list-style-type: none"> <li>✓ userid</li> <li>✓ course</li> <li>✓ module-forum</li> <li>✓ action-adddiscussion</li> </ul>           | <b>Número respuestas:</b> Número de veces que ingresa al foro e interactúa con sus respuestas a determinado tema planteado por el profesor.          |
| <ul style="list-style-type: none"> <li>✓ userid</li> <li>✓ course</li> <li>✓ module-upload</li> <li>✓ action-upload</li> </ul>                 | <b>Número de tareas:</b> Número de archivos, entiéndase por .doc, .pdf o cualquier otro formato correspondientes a tareas subidas por el estudiante. |

La tabla 2.4. Entidad que se encarga de recoger todos los registros de las acciones que los estudiantes realizan tanto en el curso como en las herramientas, que se encuentran disponibles en la plataforma del EVA.

Tabla 2.5: mdl\_twt\_utpl

| d) mdl_twt_utpl | DESCRIPCIÓN   |
|-----------------|---|
| <b>Atributo</b> |   |
| ✓ id            | Identifica al mensaje twitter   |
| ✓ user_id       | Identifica al usuario el cual hace un cuestionamiento sobre el curso, a través de uno varios mensajes twitter |
| ✓ stamp         | Indica la fecha y hora en que fue enviado el mensaje twitter  |
| ✓ body          | Es el texto del mensaje   |
| ✓ courseid      | Identifica al curso sobre el cual van dirigidos los cuestionamientos del usuario.                             |

La tabla 2.5. Esta entidad nos permite a través del user\_id y courseid conocer la cantidad de mensajes twitter que un usuario hace en determinado curso, éstos atributos son esenciales a la hora de la construcción de la consulta SQL a la base de datos del EVA.

*Tabla 2.6: mdl\_quiz\_attempts*

| e) mdl_quiz_attempts | DESCRIPCIÓN  |
|----------------------|--|
| Atributo             |  |
| ✓ id                 | Identifica al cuestionario   |
| ✓ userid             | Identifica al usuario  |
| ✓ timestart          | Tiempo en que empieza a responder el cuestionario                        |
| ✓ timefinish         | Tiempo en que termina de responder el cuestionario                       |
| ✓ attempt            | Indica los intentos que hace el estudiante por responder el cuestionario |

La tabla 2.6. Es necesaria para poder detectar de manera rápida y permite el cálculo del tiempo total de un estudiante para resolver un cuestionario previamente puesto en el curso por el profesor de la asignatura. Así como también permite saber el número de intentos que ha realizado el estudiante en el cuestionario al momento de responder al cuestionario.

## CAPÍTULO 3

*Minería de datos para determinar  
comportamientos del estudiante relacionados  
a las acciones en el EVA*

### 3.1 Descripción del escenario

La base de datos para la búsqueda de patrones, contiene información de todos los cursos, en las modalidades (Clásica y Abierta) de estudio en la UTPL del periodo correspondiente a Abril2011/Agosto2011, conviene indicar que lo que interesa es la información de los estudiantes de la modalidad Abierta la cual será el foco del desarrollo de la tesis para una representación de indicadores en un modelo de estudiante en base a los patrones de comportamiento en relación con las acciones. Para este estudio se tomará como muestra los cursos detallados en la tabla 3.1, con el objeto de tener una muestra representativa que permita identificar diferentes comportamientos de estudiantes de algunas carreras, en base a estos comportamientos, se determinaran atributos e indicadores que pueden ser representados en un modelo de estudiante para la UTPL, que considere información que describa en gran medida las acciones que los estudiantes de la UTPL realizan sobre la plataforma, es por ello que en este caso se ha creído conveniente abarcar al menos dos carreras.

*Tabla 3.1: Muestra de las carreras con sus respectivos cursos*

| CARRERA     | ÁREA              | CURSO  | CANTIDAD DE ESTUDIANTES |
|-------------|-------------------|--|-------------------------|
| Informática | Técnica           | LÓGICA DE LA PROGRAMACIÓN                                  | 205                     |
|             |                   | MATEMÁTICAS DISCRETAS                                      | 142                     |
|             |                   | INTELIGENCIA ARTIFICIAL                                    | 18                      |
|             |                   | SEMINARIO II (Desarrollo de páginas web dinámicas con php) | 23                      |
|             |                   |  | <b>Total:388</b>        |
| Abogacía    | Socio Humanística | MEDICINA LEGAL   | 119                     |
|             |                   | ÉTICA Y DERECHOS HUMANOS                                   | 156                     |
|             |                   | LEGISLACIÓN MERCANTIL MONETARIA Y BANCARIA                 | 122                     |
|             |                   |  | <b>Total:397</b>        |

### 3.1.1 Algoritmos para determinar los comportamientos del estudiante en el EVA

Una de las técnicas que se empleará en la minería para determinar el comportamiento de los estudiantes en base a las acciones que éste realiza sobre el EVA de la UTPL, es la de **clasificación**, como son los **árboles de decisión**, **reglas de decisión** y **métodos bayesianos** éstos se utilizan para el indicador de la participación del estudiante en el curso ya que según lo analizado ayudan a predecir una o más variables discretas, basándose en otros atributos del conjunto de datos, el algoritmo hace predicciones, mientras que en el **análisis de secuencias**, el algoritmo de clústeres de secuencia permite explorar los datos que contienen eventos que pueden vincularse mediante rutas o secuencias. El algoritmo encuentra las secuencias más comunes mediante la agrupación, o agrupación en clústeres (clustering), de las secuencias que son idénticas. Estos algoritmos se encuentran detallados en el capítulo I. Este último método de clústeres es utilizado para el indicador de la utilización de las herramientas analizando de manera individual cada atributo que representan las acciones de los estudiantes para de esta manera proceder a establecer los grupos y determinar cual hace más uso de las herramientas dentro de la plataforma.

Tabla 3.2: Algoritmos para experimentación

| Algoritmos  | Opción      |
|---|-------------|
| <ul style="list-style-type: none"> <li>✓ C4.5 o J48</li> <li>✓ REPTree</li> </ul>         | Trees       |
| <ul style="list-style-type: none"> <li>✓ BayesNet</li> <li>✓ NaiveBayes</li> </ul>        | Classifiers |
| <ul style="list-style-type: none"> <li>✓ JRip</li> </ul>                                  | Rules       |
| <ul style="list-style-type: none"> <li>✓ SimpleKMeans</li> <li>✓ FarthestFirst</li> </ul> | Clústeres   |

### 3.1.2 Análisis de los datos

En esta dimensión se enfocará concretamente a la interacción personal del estudiante con la plataforma y es usada para conocer **los comportamientos de los estudiantes, según las acciones que realizan en el eva**, previo a esto se realiza un **análisis de los atributos** detectados en cada curso detallados en tabla 2.1 que en realidad sirven, ya que a partir de éstos se derivan algunas variables que pueden ser incluidas las cuales son:

- ❖ número de accesos al curso.
- ❖ número de accesos a las tareas enviadas por el profesor
- ❖ número de veces que revisa un cuestionario
- ❖ número de veces que descarga un recurso.
- ❖ número de veces que el usuario revisa su perfil.
- ❖ número de veces que el usuario accede a un foro.
- ❖ número de veces que contesta un foro.
- ❖ número de veces que revisa una discusión dentro de un foro.
- ❖ número de veces que actualiza un post dentro del foro.
- ❖ número de veces que agrega un tema de discusión o debate.
- ❖ número de veces que actualiza el perfil de usuario.
- ❖ número de mensajes que envía el usuario
- ❖ número de visitas al historial de los mensajes.
- ❖ número de veces que envía o sube una tarea.
- ❖ número de mensajes (twitter).
- ❖ número de intentos en la evaluación a distancia del primer bimestre.
- ❖ tiempo que tarda en la evaluación a distancia del primer bimestre.
- ❖ número de intentos en la evaluación a distancia del segundo bimestre.
- ❖ tiempo que tarda en la evaluación a distancia del segundo bimestre.
- ❖ Tiempo promedio de acceso al curso

La interacción del usuario con el sistema es seguida a través de los logs. Los recursos (materiales) son ayudas digitales que contienen información que puede ser leída, vista o bajada para extraer información que apoye el proceso de aprendizaje, éstos permiten conocer el material que ha visto el estudiante.

### **3.1.3 Indicadores correspondientes a las acciones del estudiante con el EVA**

Un **indicador**, según algunas referencias es descrito como: “Herramientas para clarificar y definir de, forma más precisa, objetivos e impactos y son medidas verificables de cambio o resultado diseñadas para contar con un estándar contra el cual evaluar, estimar o demostrar el progreso con respecto a metas establecidas, facilitan el reparto de insumos, produciendo productos y alcanzando objetivos”<sup>7</sup>.

Entre las características que poseen los indicadores caben indicar las siguientes:

- ❖ Estar inscrito en un marco teórico o conceptual, que le permita asociarse firmemente con el evento al que el investigador pretende dar forma.
- ❖ Ser específicos, es decir, estar vinculados con los fenómenos en la educación sobre los que se pretende intervenir, para ello estos indicadores deben contar con objetivos y metas claras, para poder evaluar que tan cerca o lejos se encuentra de los mismos y proceder a una toma de decisión pertinente. Deben mostrarse especificando la meta u objetivo a que se vinculan y a lo que se pretende dar seguimiento; para cumplirlas se recomienda que los indicadores sean pocos.
- ❖ Ser explícitos, de tal forma que su nombre sea suficiente para entender si se trata de un valor absoluto o relativo, de una tasa, una razón, un índice, etc., así como a qué

---

<sup>7</sup> Organización de las Naciones Unidas (ONU). Integrated and coordinated implementation and follow-up of major. United Nations conferences and summits. Nueva York, Estados Unidos de América, 10 y 11 de mayo de 1999, p. 18. Recuperado el 1 de Junio 2011, de <http://bit.ly/qkyoxM>.

grupo de estudiantes se refieren, a quienes, que hacen mayor y menor utilidad de una herramienta así como también su participación en un curso.

- ❖ Estar disponibles para varios años, con el fin de que se pueda observar el comportamiento del fenómeno a través del tiempo. La comparabilidad permite fomentar el desarrollo educativo de acuerdo con el contenido que se les imparte a los estudiantes o las herramientas con las que se trabaja, respecto a los demás o a otros momentos.
- ❖ Los indicadores no son exclusivos de una acción específica; uno puede servir para estimar el impacto de dos o más hechos, o viceversa. Solo de manera excepcional, un indicador proveerá información suficiente para la comprensión de fenómenos tan complejos como la educación; por lo que, para tener una evaluación completa de una población, se requiere de un conjunto de indicadores que mida el desempeño de las distintas entidades y proporcione información acerca de la manera como operan para producir un efecto general.
- ❖ Ser claro, de fácil comprensión, para cada indicador debe existir una definición, fórmula de cálculo y metadatos necesarios para su mejor entendimiento, por esto es importante que el indicador sea confiable, en cuanto a su metodología de cálculo y consistente, que permita producir la misma conclusión si la medición es llevada a cabo con diferentes herramientas, por distintas personas, en similares circunstancias.
- ❖ Que la recolección de la información permita construir el mismo indicador de la misma manera y bajo condiciones similares, año tras año de modo que las comparaciones sean válidas.
- ❖ Técnicamente debe ser sólido, es decir, válido, confiable y comparable, así como factible en términos de que su medición tenga un costo razonable.
- ❖ Ser sensible a cambios en el fenómeno, tanto para mejorar como para empeorar.

Las características de los indicadores manifestadas con anterioridad sirvieron para, establecer los indicadores que corresponden a los datos sobre las acciones que los estudiantes, realizan en el entorno virtual de aprendizaje y se ha creído conveniente establecer los siguientes:

- ❖ Participación o interés en el curso.
- ❖ Utilización de herramientas.

**Participación o interés en el curso:** Este representa el interés que el estudiante tiene en determinado curso midiendo la aportación que éste haya realizado en la plataforma, datos útiles para poder sugerir sobre contenidos a los estudiantes que presenten un nivel de interés bajo, y contenidos un poco más complejos para los estudiantes que presenten un nivel de participación alto. En (Alberca, Loja, Valdiviezo, Jiménez y Agila, 2008), presentan un apartado en el que describen el *tipo de interacción del usuario*, esto ha servido como base para delimitar los niveles en cada indicador que se ha establecido para efectos de esta tesis, ya que en el trabajo que ellos realizaron muestran la interacción conformista tipificándola como *bajo*, interacción consiente como *medio* e interacción autónoma como *alto*, lo que da pie y se cree conveniente que para este trabajo, en cada uno de los indicadores que presenten valores de incidencia alto será denotado como *permanente* (P), valores de incidencia medio como *moderado* (M) y los valores de incidencia bajo como *escaso* (E). Quedando así establecidos los siguientes niveles:

- ❖ Permanente (P)
- ❖ Moderado (M)
- ❖ Escaso (E)

**Utilización de herramientas:** Aquí se trata de las acciones que el estudiante realiza sobre las herramientas tales como foros, mensajes, recursos, tareas, twitter, y cuestionario, puesto que son actividades subidas por el profesor en la plataforma para el aprendizaje

del alumno. Para este indicador primero se detalla las acciones por ejemplo para el caso de los foros tenemos algunas acciones como son: ver foro, ver discusión dentro de un foro, añadir discusión, añadir post y actualizar post, estas acciones sobre la respectiva herramienta se encuentran detallado en la tabla 3.1 . Aquí también utilizaremos los siguientes niveles:

- ✓ Permanente (P)
- ✓ Moderado (M)
- ✓ Escaso (E)

Para representar el nivel de utilización de herramientas que tiene el estudiante sobre la plataforma.

En la tabla 3.3, se presenta una descripción se las acciones que los estudiantes realizan en cada una de las herramientas, que les ofrece el curso del entorno virtual de aprendizaje.

*Tabla 3.3: Acciones en las herramientas utilizadas en los cursos*

| Herramienta         | Acción                                       |
|---------------------|--|
| <b>foros</b>        | Ver foro                                     |
|                     | Añadir post                                  |
|                     | Ver discusión dentro del foro                |
|                     | Actualizar post                              |
|                     | Añadir tema de discusión o debate            |
| <b>recursos</b>     | Descargar archivos                           |
| <b>tareas</b>       | Visualizar tareas propuestas por el profesor |
|                     | Enviar tareas                                |
| <b>twitter</b>      | Enviar mensaje twitter                       |
| <b>mensajes</b>     | Escribir mensajes                            |
|                     | Ver historial de mensajes                    |
| <b>cuestionario</b> | Responder un cuestionario                    |
|                     | Visualizar preguntas                         |

La evaluación de éstos niveles de seguimiento tiene su fundamento en la utilización de indicadores ya que los datos obtenidos permiten la acción tutorial cuando se detecta una desatención del alumno de sus actividades académicas, ya que según la participación del alumno en el entorno virtual, y con los atributos mencionados anteriormente se podrá dar soporte acertado en cuanto a contenidos en los cursos que se mencionan en la descripción del escenario sección 3.1.

### 3.1.4 Escenario de experimentación

Los algoritmos y datos que se utilizaron en las experimentaciones que se realizaron con la herramienta **WEKA**, dentro de la minería de datos se detallan a continuación:

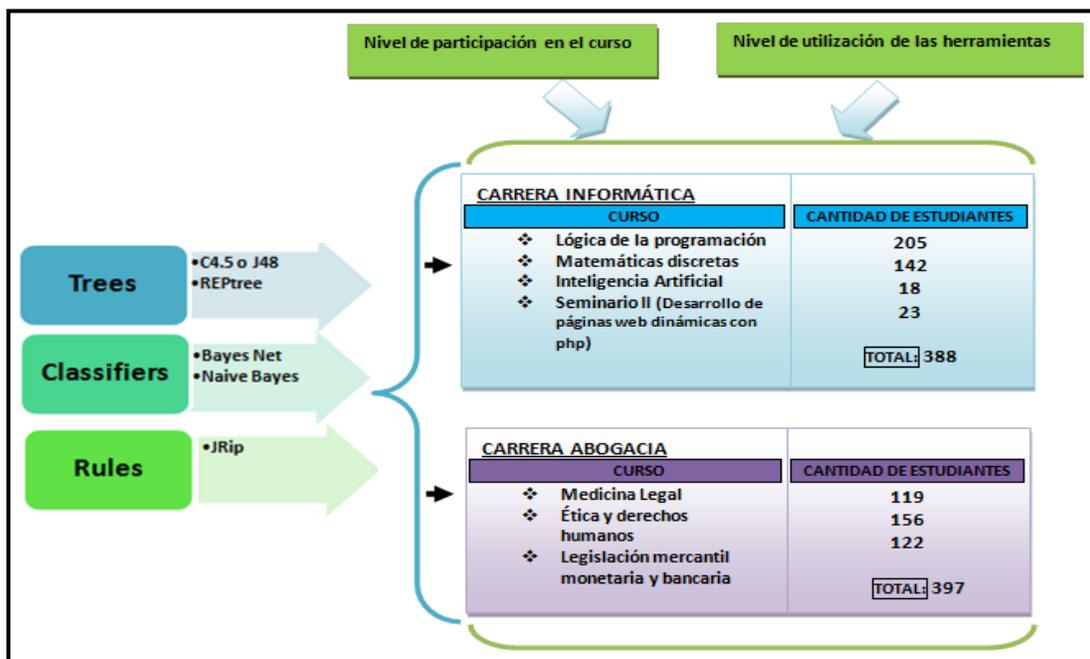


Figura 3.1: Escenario para el nivel de participación en el curso y el nivel de utilización de las herramientas por carrera.

En la figura 3.1, se puede apreciar los algoritmos C4.5 o J48 (Trees), Bayes Net y Naive Bayes (Classifiers o métodos bayesianos) y JRip (Rules o reglas de decisión) para la

experimentación, los cuales forman parte de la **técnica de clasificación**, estos serán aplicados a cada una de las carreras para encontrar dos indicadores: el *nivel de participación en el curso* y el *nivel de utilización de las herramientas*, en el caso de informática se cuenta con 388 instancias o cantidad de estudiantes mientras que en abogacía 397 instancias, a partir de la experimentación se seleccionará el algoritmo que presente los mejores resultados para ser sujetos a análisis e interpretación y posteriormente proceder a realizar comparaciones entre los resultados obtenidos de ambas carreras ver punto 3.6.3, previo a esto se realizó una descripción de los algoritmos que se utilizarán (en el punto 3.1.1), y un análisis de los datos de cada curso de ahí se obtuvieron algunas variables ver (en el punto 3.1.2). Las variables que se presentan son los datos a extraer de cada curso, mediante la utilización de sentencias SQL ver en anexo B.

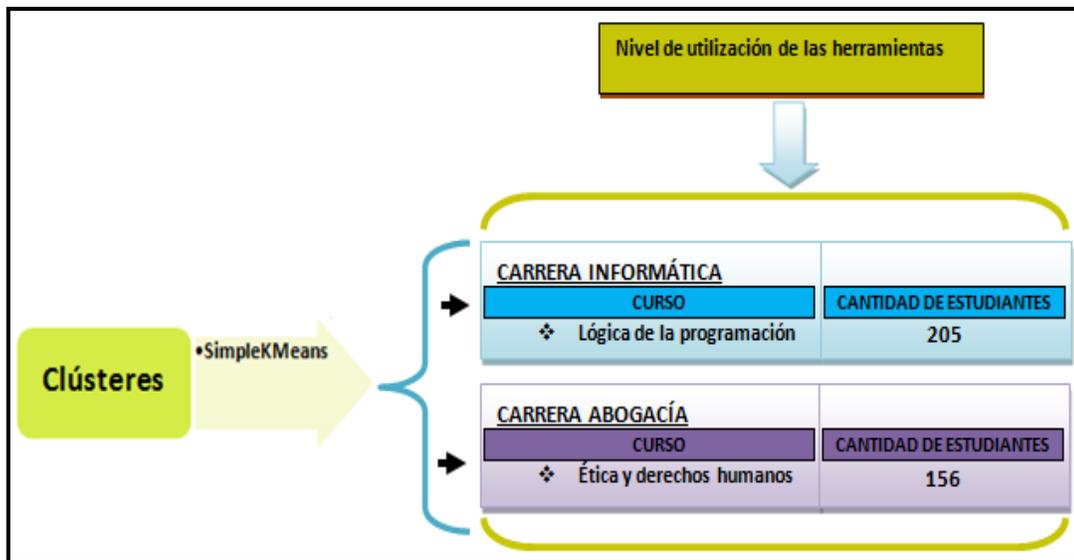


Figura 3.2: Escenario para el nivel de utilización de las herramientas por curso

En la figura 3.2, se presenta un esquema de la aplicación del algoritmo SimpleKMeans (Clústeres) para la experimentación, este algoritmo fue aplicado a un curso de cada carrera, para encontrar el indicador del *nivel de utilización de las herramientas* por curso, ya no por carrera como el escenario de la figura 3.1, esta vez se selecciona el que tenga la mayor cantidad

de estudiantes o instancias, ya que al contar con un número de instancias considerable, los resultados obtenidos son más confiables, los cursos resultantes son: *Lógica de la programación* de la carrera de Informática con 205 instancias y, *ética y derechos humanos de Abogacía* con 156 instancias.

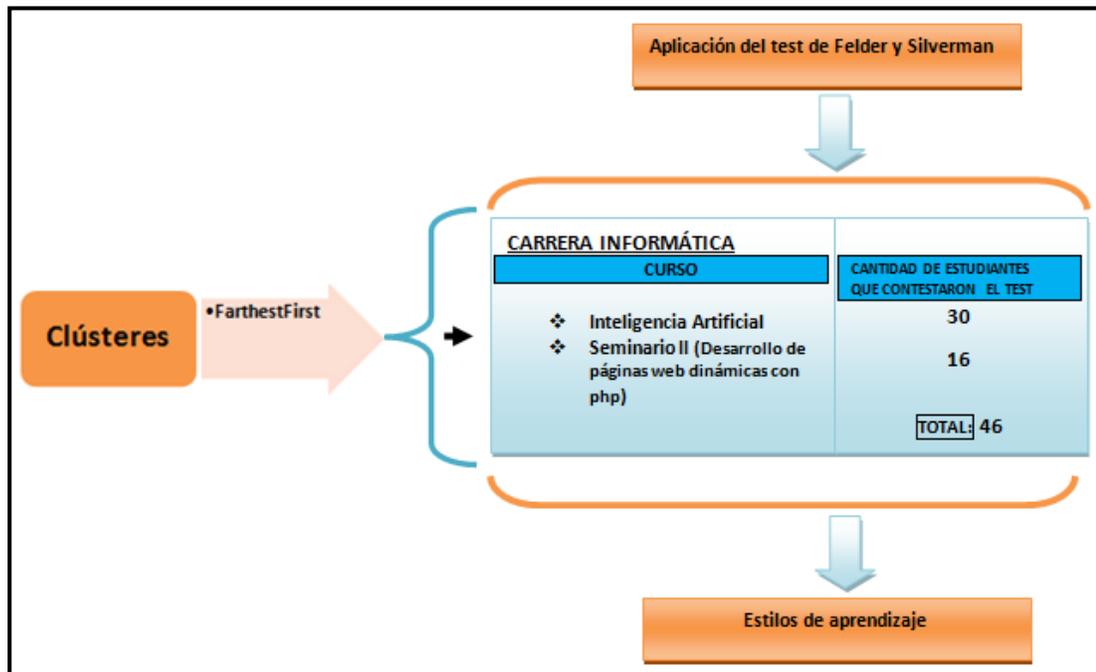


Figura 3.3: Escenario para encontrar el estilo de aprendizaje dominante

En la figura 3.3, se puede observar el algoritmo FarthestFirst (Clústeres), el cual se aplicará a los cursos de los últimos ciclos de la carrera de informática: *Inteligencia Artificial* y *Seminario II (Desarrollo de páginas web dinámicas con php)* para encontrar el estilo de aprendizaje predominante y realizar un análisis de cómo aprenden los estudiantes en esta carrera, previa a la aplicación del algoritmo se habilita a manera de enlace, el test de Felder y Silverman a los estudiantes en cada curso quedando como resultado un total de 46 estudiantes que dieron contestación al test.

## 3.2 Limpieza y transformación de datos

La recopilación de datos debe ir acompañada de una limpieza e integración de los mismos, para que estos datos estén en condiciones para su análisis. Los beneficios del análisis y de la extracción de conocimiento a partir de datos dependen, en gran medida, de la calidad de los datos recopilados. Además, generalmente, debido a las características propias de las técnicas de minería de datos, es necesario realizar una transformación de los datos para obtener una “materia prima” que sea adecuada para el propósito concreto y las técnicas que se quieren emplear. En definitiva, el éxito de un proceso de minería de datos depende, no solo de tener todos los datos necesarios (una buena recopilación), sino de que estén íntegros, completos y consistentes (una buena limpieza e integración) (Hernández, Ramírez y Ferri, 2008).

En esta fase se ha estimado conveniente la transformación del tiempo de la tabla `mdl_quiz_attempts` descrita en el Capítulo 2, ya que originalmente en la base de datos se encontraba en un formato no legible para humanos esta es `timestamp`, haciendo uso de la función `select from_unixtime()`; esto nos facilita las unidades para saber el tiempo que un estudiante tarda en la resolución de un cuestionario. Consolidando estos datos apropiadamente para continuar con la siguiente fase. Para la parte de integración de las distintas fuentes de datos registradas primero en Excel, se procedió a crear tablas en una base de datos de las asignaturas aquí recopilaremos todos los atributos que nos serán de utilidad.

Con la información obtenida a través de las consultas SQL que se muestran en el anexo B figura B.1 para recopilar todas las acciones que se realizaron en determinado curso, entre las actividades de pre procesamiento se aplicó limpieza de los datos en el momento de ejecutar la consulta ya que en el caso de que hubieran existido estudiantes que constaban como participantes del curso pero sin embargo nunca ni siquiera ingresaron al curso en el EVA, por ende éstos no realizaron ninguna acción y no fueron tomados en cuenta, también se eliminaron algunas inconsistencias a través de correcciones manuales, tales como la existencia de

espaciados innecesarios. Adicional a esto se realizó otra consulta que está en el anexo B figura B.2, para las acciones en los cuestionarios, tales como el número de intentos y tiempo que emplea el estudiante para rendir el módulo antes mencionado. Para efecto del registro de cada una de las acciones, se lo colocó en una hoja de cálculo Excel, posteriormente se almaceno en la base de datos *minería acciones* para luego ser cargado en WEKA.

### 3.3 Discretización

La discretización, es la conversión de un valor numérico en un valor nominal ordenado el cual representa un intervalo o “bin”, cuyo proceso consiste en la búsqueda de los intervalos más adecuados para cada dominio numérico, para esto lo primero que se realizó fue conectar desde WEKA<sup>8</sup> la base de datos *minería acciones* donde se recopiló información necesaria para la extracción del conocimiento, base que antes fue alojada en el manejador de base de datos SQLyog8.05 mediante sentencias SQL mismas que se guardaron en un .txt o archivo de texto plano.

#### 3.3.1 Discretización para los valores de los atributos de la carrera de *informática*.

Con la ayuda de una consulta figura 3.4 se extrajo a la aplicación los atributos que servirán para determinar la participación en el curso, mismos atributos que se procederán a discretizar desde WEKA.

---

<sup>8</sup> WEKA (Waikato Environment for Knowledge Analysis - Entorno para Análisis del Conocimiento de la Universidad de Waikato) <http://www.cs.waikato.ac.nz/ml/weka/>

```
1 SELECT DISTINCT
2 informatica.userid,
3 informatica.num_accesos_curso,
4 informatica.tavg_acceso_curso/num_accesos_curso AS tavg_acceso_curso,
5 informatica.num_accesos_tareas,
6 informatica.num_veces_sube_tarea,
7 informatica.num_veces_descarga_recurso,
8 informatica.num_veces_revisa_cuestionario,
9 informatica.num_intevaladpb,
10 informatica.tie_evaladpb/num_intevaladpb AS tie_evaladpb,
11 informatica.num_intevaladsb,
12 informatica.tie_evaladsb/num_intevaladsb AS tie_evaladsb,
13 informatica.num_msn_escritos,
14 informatica.num_veces_revisa_historial_msn,
15 informatica.num_veces_revisa_perfil,
16 informatica.num_accesos_foro,
17 informatica.num_veces_contesta_foro,
18 informatica.`num_veces_revisa_discusión_foro`,
19 informatica.num_veces_actualiza_contest_foro,
20 informatica.`num_veces_agrega_tema_discusión`,
21 informatica.num_veces_actualiza_perfil,
22 informatica.num_msn_twitter,
23 informatica.`participación_curso`
24 FROM
25 informatica
26 GROUP BY
27 informatica.userid,
28 informatica.num_accesos_curso,
29 informatica.num_accesos_tareas,
30 informatica.num_veces_sube_tarea,
31 informatica.num_veces_descarga_recurso,
32 informatica.num_veces_revisa_cuestionario,
33 informatica.num_intevaladpb,
34 informatica.num_intevaladsb,
35 informatica.num_accesos_foro,
```

Figura 3.4: Consulta SQL para extraer atributos de la participación en el curso (informática).

Luego se utilizó un filtro, no supervisado "*weka.filters.unsupervised.attribute.Discretize*" desde weka para con ayuda de esta herramienta optimizar el número de intervalos, en lo referente a la configuración en el parámetro bins se le colocó un valor de 3 que corresponden a los intervalos de los niveles *escaso*, *moderado* y *permanente*, para los atributos predictivos sobre la participación de un estudiante en el curso según las acciones que realiza ver en la tabla 3.3, cabe indicar que el mismo procedimiento se realizó para la utilización de las herramientas.

**Tabla 3.4: Discretización para el nivel de participación en el curso (informática)**

| <b>388 INSTANCIAS</b><br><b>ATRIBUTOS</b>   | <b>INTERVALOS</b> |                    |                      |
|---|-------------------|--------------------|----------------------|
|   | <b>Escaso(E)</b>  | <b>Moderado(M)</b> | <b>Permanente(P)</b> |
| <i>num_accesos_curso</i>  | -inf-75           | 75-150             | 150-inf              |
| <i>tavg_acceso_curso</i> (tiempo promedio de acceso al curso)                       | -inf-27355        | 27355-54690        | 54960-inf            |
| <i>num_accesos_tareas</i>   | -inf-23           | 23-46              | 46-inf               |
| <i>numveces_sube_tarea</i>  | -inf-7            | 7-15               | 15-inf               |
| <i>num_veces_descarga_recurso</i>   | -inf-194          | 194-388            | 388-inf              |
| <i>num_veces_revisa_cuestionario</i>  | -inf-24           | 24-48              | 48-inf               |
| <i>num_intevaldpb</i> (número de intentos, evaluación a distancia primer bimestre)  | -inf-1            | 1-2                | 2-inf                |
| <i>tie_evaladpb</i> (tiempo, evaluación a distancia primer bimestre)                | -inf-33333        | 33333-66666        | 66666-inf            |
| <i>num_intevaldsb</i> (número de intentos, evaluación a distancia segundo bimestre) | -inf-1            | 1-2                | 2-inf                |
| <i>tie_evaladsb</i> (tiempo, evaluación a distancia segundo bimestre)               | -inf-33375        | 33375-66687        | 66687-inf            |
| <i>num_msn_escritos</i>   | -inf-13           | 13-27              | 27-inf               |
| <i>num_veces_revisa_historial_msn</i>   | -inf-30           | 30-60              | 60-inf               |
| <i>num_veces_revisa_perfil</i>  | -inf-37           | 37-74              | 74-inf               |
| <i>num_veces_actualiza_perfil</i>   | -inf-2            | 2-4                | 4-inf                |
| <i>num_accesos_foro</i>   | -inf-63           | 63-127             | 127-inf              |
| <i>num_veces_contesta_foro</i>  | -inf-10           | 10-20              | 20-inf               |
| <i>num_veces_revisa_discusión_foro</i>  | -inf-53           | 53-107             | 107-inf              |
| <i>num_veces_actualiza_contest_foro</i>   | -inf-4            | 4-8                | 8-inf                |
| <i>num_veces_agrega_tema_discusión</i>  | -inf-2            | 2-4                | 4-inf                |
| <i>num_msn_twitter</i>  | -inf-7            | 7-14               | 14-inf               |

En la tabla 3.4, se observa los valores de cada uno de los atributos con su correspondiente intervalo que previamente fue realizado con las configuraciones pertinentes en la herramienta weka, por ejemplo en el caso del atributo (*num\_accesos\_curso*), los valores que se encuentren en el intervalo desde el menos infinito a 75 (-inf-75) concierne al *nivel de participación en el curso* (escaso), los valores ubicados en el intervalo entre 75 hasta 150 (75-150) pertenece al *nivel de participación en el curso* (moderado), y los valores que van en el intervalo de 150 al infinito (150-inf) pertenece al *nivel de participación en el curso* (permanente).

Ahora bien una vez que ya se tiene determinado cuales van a ser los intervalos para cada atributo se procede a través de WEKA a evaluar la participación del estudiante en el curso, según (Mladen, 2008) denota como una técnica de minería de datos en la clasificación, a las *técnicas basadas en conteo de frecuencias*, que se apoyan en contar la frecuencia con la que dos o más sucesos se presentan, lo que ha permitido incluir y dejar establecido que en este trabajo de tesis se ha creído conveniente tomar en cuenta la *moda*<sup>9</sup> en cada atributo y por cada fila de nuestro registro de datos como se muestra en la figura 3.5 se asigna el nivel ya sea Permanente (P), Moderado (M) y Escaso (E). Éstos pasos se realizan por cada curso los que se detallan aquí son los referentes a la carrera de *Informática*.

| Relation: QueryResult | 19: num_veces_agrega_tema_discusión<br>Numeric | 20: num_veces_actualiza_perfil<br>Numeric | 21: num_msn_twitter<br>Numeric | 22: participación_curso<br>Nominal |
|-----------------------|--|---|--------------------------------|------------------------------------|
| 6.0                   | 2.0  | 4.0                                       | 3.0                            | M                                  |
| 0.0                   | 0.0  | 0.0                                       | 0.0                            | E                                  |
| 0.0                   | 0.0  | 0.0                                       | 0.0                            |                                    |
| 0.0                   | 0.0  | 0.0                                       | 8.0                            | P                                  |
| 0.0                   | 2.0  | 1.0                                       | 2.0                            | M                                  |
| 0.0                   | 2.0  | 0.0                                       | 0.0                            | E                                  |

Figura 3.5: Asignación de la clase para el indicador del nivel de participación (informática)

<sup>9</sup>Moda, es el valor que cuenta con una mayor frecuencia en una distribución de datos. Se puede hallar la moda para variables cualitativas y cuantitativas. <http://bit.ly/pkxuQ8>

Continuamos construyendo una consulta SQL figura 3.6 para extraer a la aplicación WEKA los atributos que servirán para determinar el nivel de utilización de las herramientas que hacen los estudiantes en la plataforma.

```
1 SELECT DISTINCT
2 informatica.userid,
3 informatica.view_forum_forum,
4 informatica.add_post_forum,
5 informatica.view_discussion_forum,
6 informatica.update_postforum,
7 informatica.add_discussion,
8 informatica.view_resource,
9 informatica.view_assignment,
10 informatica.upload_upload,
11 informatica.ms_twt,
12 informatica.write_message,
13 informatica.history_message,
14 informatica.view_quiz,
15 informatica.num_intevaladpb,
16 informatica.tie_evaladpb/num_intevaladpb AS tie_evaladpb,
17 informatica.num_intevaladsb,
18 informatica.tie_evaladsb/num_intevaladsb AS tie_evaladsb,
19 informatica.`utilización_herramientas`
20 FROM
21 informatica
22 GROUP BY
23 informatica.userid,
24 informatica.view_forum_forum,
25 informatica.add_post_forum,
26 informatica.view_discussion_forum,
27 informatica.update_postforum,
28 informatica.add_discussion,
29 informatica.view_resource,
30 informatica.view_assignment,
31 informatica.upload_upload,
32 informatica.ms_twt,
33 informatica.write_message,
34 informatica.history_message,
35 informatica.view_quiz,
36 informatica.num_intevaladpb,
```

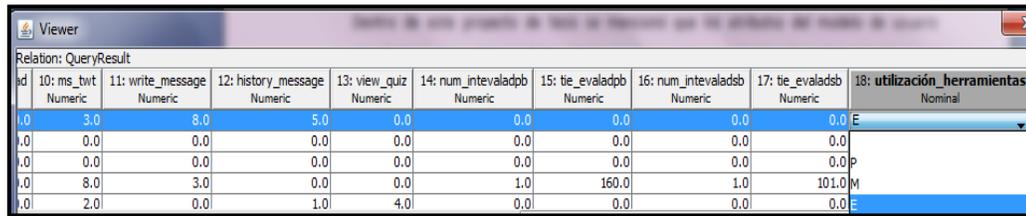
Figura 3.6: Consulta SQL para extraer atributos de la utilización de herramientas (informática).

De igual forma que se hizo la discretización para el indicador de la participación en el curso ahora se realiza la discretización para el indicador de la utilización de las herramientas por parte de los estudiantes ver tabla 3.4

*Tabla 3.5: Discretización para el nivel de utilización de herramientas (informática)*

| 388 INSTANCIAS<br>ATRIBUTOS   | INTERVALOS |             |               |
|---|------------|-------------|---------------|
|   | Escaso(E)  | Moderado(M) | Permanente(P) |
| <i>view_forum_forum</i>   | -inf-63    | 63-127      | 127-inf       |
| <i>add_post_forum</i>   | -inf-10    | 10-20       | 20-inf        |
| <i>view_discussion_forum</i>  | -inf-53    | 53-107      | 107-inf       |
| <i>update_postforum</i>   | -inf-4     | 4-8         | 8-inf         |
| <i>add_discussion</i>   | -inf-2     | 2-4         | 4-inf         |
| <i>view_resource</i>  | -inf-194   | 194-388     | 388-inf       |
| <i>view_assignment</i>  | -inf-23    | 23-46       | 46-inf        |
| <i>upload_upload</i>  | -inf-7     | 7-15        | 15-inf        |
| <i>ms_twitter</i>   | -inf-7     | 7-14        | 14-inf        |
| <i>write_message</i>  | -inf-13    | 13-27       | 27-inf        |
| <i>history_message</i>  | -inf-30    | 30-60       | 60-inf        |
| <i>view_quiz</i>  | -inf-24    | 24-48       | 48-inf        |
| <i>num_intevaldpb</i> (número de intentos, evaluación a distancia primer bimestre)  | -inf-1     | 1-2         | 2-inf         |
| <i>tie_evaladpb</i> (tiempo, evaluación a distancia primer bimestre)                | -inf-33333 | 33333-66666 | 66666-inf     |
| <i>num_intevaldsb</i> (número de intentos, evaluación a distancia segundo bimestre) | -inf-1     | 1-2         | 2-inf         |
| <i>tie_evaladsb</i> (tiempo, evaluación a distancia segundo bimestre)               | -inf-33375 | 33375-66687 | 66687-inf     |

Una vez lista la discretización se continúa con la asignación de los niveles figura 3.7 teniendo en cuenta que primará el que tenga mayor concurrencia en los valores de los atributos por cada fila, para esto se realiza la tabulación a parte en un archivo de Excel.



| id  | 10: ms_twt<br>Numeric | 11: write_message<br>Numeric | 12: history_message<br>Numeric | 13: view_quiz<br>Numeric | 14: num_intevaladpb<br>Numeric | 15: tie_evaladpb<br>Numeric | 16: num_intevaladsb<br>Numeric | 17: tie_evaladsb<br>Numeric | 18: utilización_herramientas<br>Nominal |
|-----|-----------------------|------------------------------|--------------------------------|--------------------------|--------------------------------|-----------------------------|--------------------------------|-----------------------------|---|
| 1.0 | 3.0                   | 8.0                          | 5.0                            | 0.0                      | 0.0                            | 0.0                         | 0.0                            | 0.0                         | E                                       |
| 1.0 | 0.0                   | 0.0                          | 0.0                            | 0.0                      | 0.0                            | 0.0                         | 0.0                            | 0.0                         |   |
| 1.0 | 0.0                   | 0.0                          | 0.0                            | 0.0                      | 0.0                            | 0.0                         | 0.0                            | 0.0                         | P                                       |
| 1.0 | 8.0                   | 3.0                          | 0.0                            | 0.0                      | 1.0                            | 160.0                       | 1.0                            | 101.0                       | M                                       |
| 1.0 | 2.0                   | 0.0                          | 1.0                            | 4.0                      | 0.0                            | 0.0                         | 0.0                            | 0.0                         | E                                       |

Figura 3.7: Asignación de la clase para el indicador de la utilización de la herramienta (informática).

### 3.3.2 Discretización para los valores de los atributos de la carrera de *abogacía*.

Con la ejecución de la consulta de la figura 3.8 y previa la conexión a la base de datos desde WEKA se inicia el primer paso para proceder a realizar la discretización para los valores del nivel de participación o interés de los estudiantes en el curso, cabe indicar que antes ya se realizó la discretización en el caso de los valores que competen a los atributos de la carrera de *informática* ver en el punto 3.3.1 pues se siguió exactamente los mismos pasos.

```

1 SELECT DISTINCT
2 abogacia.userid,
3 abogacia.num_accesos_curso,
4 abogacia.tavg_acceso_curso/num_accesos_curso AS tavg_acceso_curso,
5 abogacia.num_accesos_tareas,
6 abogacia.num_veces_sube_tarea,
7 abogacia.num_veces_descarga_recurso,
8 abogacia.num_veces_revisa_cuestionario,
9 abogacia.num_intevaladpb,
10 abogacia.tie_evaladpb/num_intevaladpb AS tie_evaladpb,
11 abogacia.num_intevaladsb,
12 abogacia.tie_evaladsb/num_intevaladsb AS tie_evaladsb,
13 abogacia.num_msn_escritos,
14 abogacia.num_veces_revisa_historial_msn,
15 abogacia.num_veces_revisa_perfil,
16 abogacia.num_accesos_foro,
17 abogacia.num_veces_contesta_foro,
18 abogacia.'num_veces_revisa_discusión_foro',
19 abogacia.'num_veces_actualiza_contest_foro',
20 abogacia.'num_veces_agrega_tema_discusión',
21 abogacia.num_veces_actualiza_perfil,
22 abogacia.num_msn_twitter,
23 abogacia.'participación_curso'
24 FROM
25 abogacia
26 GROUP BY
27 abogacia.userid,
28 abogacia.num_accesos_curso,
29 abogacia.num_accesos_tareas,
30 abogacia.num_veces_sube_tarea,
31 abogacia.num_veces_descarga_recurso,
32 abogacia.num_veces_revisa_cuestionario,
33 abogacia.num_intevaladpb,
34 abogacia.num_intevaladsb,
35 abogacia.num_accesos_foro,
36 abogacia.num_veces_contesta_foro,

```

Figura 3.8: Consulta SQL para extraer atributos de la participación en el curso (abogacía).

**Tabla 3.6: Discretización para el nivel de participación en el curso (abogacía)**

| 397 INSTANCIAS<br>ATRIBUTOS   | INTERVALOS |             |               |
|---|------------|-------------|---------------|
|   | Escaso(E)  | Moderado(M) | Permanente(P) |
| <i>num_accesos_curso</i>  | -inf-66    | 66-131      | 131-inf       |
| <i>tavg_acceso_curso (seg)</i>  | -inf-29711 | 29711-59288 | 59288-inf     |
| <i>num_accesos_tareas</i>   | -inf-4     | 4-8         | 8-inf         |
| <i>numveces_sube_tarea</i>  | -inf-1     | 1-2         | 2-inf         |
| <i>num_veces_descarga_recurso</i>   | -inf-30    | 30-60       | 60-inf        |
| <i>num_veces_revisa_cuestionario</i>  | -inf-4     | 4-8         | 8-inf         |
| <i>num_intevaldpb</i> (número de intentos, evaluación a distancia primer bimestre)  | -inf-1     | 1-2         | 2-inf         |
| <i>tie_evaladpb</i> (tiempo, evaluación a distancia primer bimestre)                | -inf-24047 | 24047-48044 | 48044-inf     |
| <i>num_intevaldsb</i> (número de intentos, evaluación a distancia segundo bimestre) | -inf-1     | 1-2         | 2-inf         |
| <i>tie_evaladsb</i> (tiempo, evaluación a distancia segundo bimestre)               | -inf-18058 | 18058-36043 | 36043-inf     |
| <i>num_msn_escritos</i>   | -inf-33    | 33-66       | 66-inf        |
| <i>num_veces_revisa_historial_msn</i>   | -inf-31    | 31-63       | 63-inf        |
| <i>num_veces_revisa_perfil</i>  | -inf-6     | 6-13        | 13-inf        |
| <i>num_veces_actualiza_perfil</i>   | -inf-1     | 1-3         | 3-inf         |
| <i>num_accesos_foro</i>   | -inf-9     | 9-18        | 18-inf        |
| <i>num_veces_contesta_foro</i>  | -inf-1     | 1-2         | 2-inf         |
| <i>num_veces_revisa_discusión_foro</i>  | -inf-2     | 2-5         | 5-inf         |
| <i>num_msn_twitter</i>  | -inf-1     | 1-2         | 2-inf         |

Esta vez para la discretización del nivel de participación o interés, se encontró una particularidad que difiere con la carrera de *informática*, pues se tiene la ausencia de los atributos *num\_veces\_actualiza\_contesta\_foro* y *núm\_veces\_agrega\_tema\_discusión*, por la razón de que los estudiantes de abogacía presentaron valores de cero, en estas acciones en foros, y no es conveniente ser sometidos a discretización ya que entorpece el descubrimiento de conocimiento convirtiéndose en ruido que no será útil para predecir el indicador de interés.

Apoyándonos en WEKA se escogió la opción del filtro no supervisado, que es quien se encarga de proveernos los valores de los intervalos para cada atributo, previa la conexión a la base de datos y ejecutando la consulta de la figura 3.8.

| perfil | 15: num_accesos_foro<br>Numeric | 16: num_veces_contesta_foro<br>Numeric | 17: num_veces_revisa_discusión_foro<br>Numeric | 18: num_veces_actualiza_perfil<br>Numeric | 19: num_msn_twitter<br>Numeric | 20: participación_curso<br>Nominal |
|--------|---------------------------------|--|--|---|--------------------------------|------------------------------------|
| 0.0    | 1.0                             | 0.0                                    | 0.0  | 0.0                                       | 0.0                            | E                                  |
| 0.0    | 0.0                             | 0.0                                    | 0.0  | 0.0                                       | 0.0                            | E                                  |
| 0.0    | 0.0                             | 0.0                                    | 0.0  | 0.0                                       | 0.0                            | E                                  |
| 3.0    | 11.0                            | 2.0                                    | 4.0  | 0.0                                       | 3.0                            | M                                  |
| 0.0    | 0.0                             | 0.0                                    | 0.0  | 0.0                                       | 0.0                            | E                                  |

Figura 3.9: Asignación de la clase para el indicador del nivel de participación en el curso (abogacía)

En la figura 3.9 se muestra la asignación del nivel de participación previa una tabulación de la moda para cada atributo y por cada fila de los datos con los que se cuenta, además se puede apreciar cuando colocamos el nivel que le corresponde teniendo presente que existen tres: Permanente (P), Moderado (M) y Escaso (E) que con anterioridad fueron descritos y establecidos en la sección 3.1.3.

```

1  SELECT DISTINCT
2  abogacia.userid,
3  abogacia.view_forum_forum,
4  abogacia.add_post_forum,
5  abogacia.view_discussion_forum,
6  abogacia.update_postforum,
7  abogacia.add_discussion,
8  abogacia.view_resource,
9  abogacia.view_assignment,
10 abogacia.upload_upload,
11 abogacia.ms_twt,
12 abogacia.write_message,
13 abogacia.history_message,
14 abogacia.view_quiz,
15 abogacia.num_intevaladpb,
16 abogacia.tie_evaladpb/num_intevaladpb AS tie_evaladpb,
17 abogacia.num_intevaladsb,
18 abogacia.tie_evaladsb/num_intevaladsb AS tie_evaladsb,
19 abogacia.`utilización_herramientas`
20 FROM
21 abogacia
22 GROUP BY
23 abogacia.userid,
24 abogacia.view_forum_forum,
25 abogacia.add_post_forum,
26 abogacia.view_discussion_forum,
27 abogacia.update_postforum,
28 abogacia.add_discussion,
29 abogacia.view_resource,
30 abogacia.view_assignment,
31 abogacia.upload_upload,
32 abogacia.ms_twt,
33 abogacia.write_message,
34 abogacia.history_message,
35 abogacia.view_quiz,
36 abogacia.num_intevaladpb,

```

Figura 3.10: Consulta SQL para extraer atributos de la utilización de herramientas (abogacía).

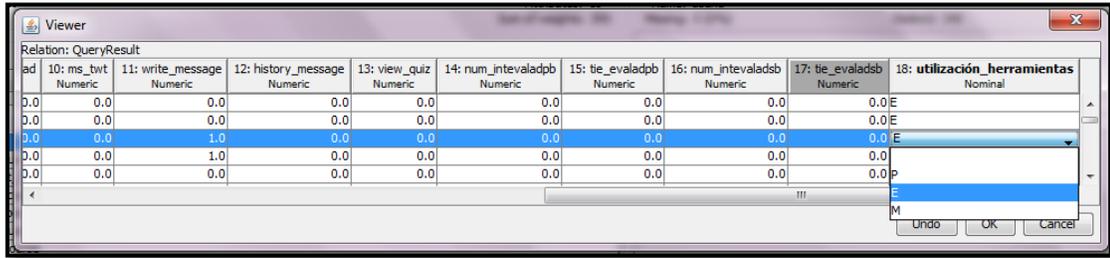
El siguiente paso es extraer a la aplicación WEKA, con ayuda de una consulta SQL figura 3.10 los atributos que nos servirán para determinar el nivel de utilización de las herramientas que hacen los estudiantes de la carrera de abogacía en la plataforma.

Ahora en la tabla 3.7 se muestran los resultados de la discretización para los valores de los atributos que, intervienen en el indicador de la utilización de las herramientas por parte de los estudiantes, se cree conveniente aclarar que esta vez no se tomo en cuenta los atributos, *update\_postforum* y *add\_discussion* por la razón de que los estudiantes no presentaron ninguna acción en éstos atributos caso contrario a la carrera de *informática*, por lo que no se estima adecuado incluirlos ya que no contribuyen con ningún aporte al proceso de KDD.

**Tabla 3.7: Discretización para el nivel de utilización de herramientas (abogacía)**

| 397 INSTANCIAS<br>ATRIBUTOS   | INTERVALOS |             |               |
|---|------------|-------------|---------------|
|   | Escaso(E)  | Moderado(M) | Permanente(P) |
| <i>view_forum_forum</i>   | -inf-9     | 9-18        | 18-inf        |
| <i>add_post_forum</i>   | -inf-1     | 1-2         | 2-inf         |
| <i>view_discussion_forum</i>  | -inf-2     | 2-5         | 5-inf         |
| <i>view_resource</i>  | -inf-30    | 30-60       | 60-inf        |
| <i>view_assignment</i>  | -inf-4     | 4-8         | 8-inf         |
| <i>upload_upload</i>  | -inf-1     | 1-2         | 2-inf         |
| <i>ms_twitter</i>   | -inf-1     | 1-2         | 2-inf         |
| <i>write_message</i>  | -inf-33    | 33-66       | 66-inf        |
| <i>history_message</i>  | -inf-31    | 31-63       | 63-inf        |
| <i>view_quiz</i>  | -inf-4     | 4-8         | 8-inf         |
| <i>num_intevaldpb</i> (número de intentos, evaluación a distancia primer bimestre)  | -inf-1     | 1-2         | 2-inf         |
| <i>tie_evaladpb</i> (tiempo, evaluación a distancia primer bimestre)                | -inf-24047 | 24047-48044 | 48044-inf     |
| <i>num_intevaldsb</i> (número de intentos, evaluación a distancia segundo bimestre) | -inf-1     | 1-2         | 2-inf         |
| <i>tie_evaladsb</i> (tiempo, evaluación a distancia segundo bimestre)               | -inf-18058 | 18058-36043 | 36043-inf     |

Luego del paso de la discretización se continúa con la asignación de los niveles de utilización de la herramienta figura 3.11 teniendo claro para realizarla, que la ubicación de cada nivel depende de la mayor concurrencia que tenga en los valores de los atributos por cada fila, para esto se realiza con anticipación la tabulación a parte en un archivo de Excel.



| ad  | 10: ms_tweet<br>Numeric | 11: write_message<br>Numeric | 12: history_message<br>Numeric | 13: view_quiz<br>Numeric | 14: num_intevaladpb<br>Numeric | 15: tie_evaladpb<br>Numeric | 16: num_intevaladsb<br>Numeric | 17: tie_evaladsb<br>Numeric | 18: utilización_herramientas<br>Nominal |
|-----|-------------------------|------------------------------|--------------------------------|--------------------------|--------------------------------|-----------------------------|--------------------------------|-----------------------------|---|
| 0.0 | 0.0                     | 0.0                          | 0.0                            | 0.0                      | 0.0                            | 0.0                         | 0.0                            | 0.0                         | E                                       |
| 0.0 | 0.0                     | 0.0                          | 0.0                            | 0.0                      | 0.0                            | 0.0                         | 0.0                            | 0.0                         | E                                       |
| 0.0 | 0.0                     | 1.0                          | 0.0                            | 0.0                      | 0.0                            | 0.0                         | 0.0                            | 0.0                         | E                                       |
| 0.0 | 0.0                     | 1.0                          | 0.0                            | 0.0                      | 0.0                            | 0.0                         | 0.0                            | 0.0                         | P                                       |
| 0.0 | 0.0                     | 0.0                          | 0.0                            | 0.0                      | 0.0                            | 0.0                         | 0.0                            | 0.0                         | P                                       |

Figura 3.11: Asignación de la clase para el indicador de la utilización de la herramienta (abogacía)

### 3.4 Minería de datos

En la minería de datos, es necesario tener los datos, ya generados como atributos, además de que en esta fase se evaluarán técnicas y algoritmos de aprendizaje automático para elegir cuál será el más apropiado para los datos que se ha seleccionado. El proceso de la minería será efectuado a través de la herramienta, de minería de datos WEKA.

Dentro de este proyecto de tesis se mencionó que los atributos a identificar del modelo de estudiante atenderán a un modelo predictivo de las instancias, y para este tipo de modelo se optará por utilizar las tareas que son clasificación y análisis de secuencias.

Una vez que se ha configurado en WEKA correctamente la moda del nivel de participación en el curso, guiándonos en los intervalos que con la ayuda de la misma herramienta se obtuvieron y para evitar posibles confusiones primero se tabuló dicha asignación en una hoja de cálculo para luego proceder a situar el nivel de participación en el curso según el número mayor de frecuencia o el mayor número de veces que se repite la clase E, M y P, ese será el valor cualitativo a asignar, ver en figura 3.9.

Terminado esto se procedió a hacer uso de los algoritmos que posee WEKA y se guardó los modelos generados por cada experiencia de algoritmo utilizada ya, que luego las matrices de confusión también serían sujetas a análisis. La técnica de evaluación utilizada fue la de cross-validation (validación cruzada) esta consiste en que: dado un número  $n$  se divide los datos en  $n$

partes y, por cada parte, se construye el clasificador con las  $n-1$  partes restantes y se prueba con esa parte. Así tenemos que por cada una de las  $n$  particiones para construir un modelo con todos los datos se obtienen sus ratios de error y precisión, promediando los  $n$  ratios de error (Hernández, Ramírez, y Ferri, 2008).

### 3.5 Experimentación y resultados con algoritmos de clasificación

Para proceder con la experimentación, se la realiza con la aplicación de cada uno de los algoritmos para las dos carreras, las cuales son Informática y Abogacía los algoritmos son aplicados para los indicadores del nivel de participación en el curso y la utilización de las herramientas, para los cuales se tomó los atributos de las acciones que realizan los estudiantes sobre la plataforma ver en la tabla 3.3, para determinar cuál de los algoritmos resultó más eficiente para el indicador a evaluar, se lo hace bajo el parámetro de la razón de precisión, que es el número de clasificaciones correctas que nos arroja en los resultados del algoritmo.

#### 3.5.1 Resultados obtenidos para la carrera de *informática*

*Tabla 3.8: Resultados obtenidos para el nivel de participación en el curso (informática)*

| INFORMATICA   |            |                                   |                                       |   |                      |
|---|------------|-----------------------------------|---------------------------------------|---|----------------------|
| NIVEL DE PARTICIPACION EN EL CURSO  | Algoritmo  | Tiempo en generar el modelo (seg) | Instancias clasificadas correctamente | Instancias clasificadas incorrectamente | Error absoluto Medio |
|   | J48        | 0.02                              | 377- (97.1649%)                       | 11 - (2.8351 %)                         | 0.0311               |
|   | REPTree    | 0                                 | 379- (97.6804%)                       | 9 - (2.3196%)                           | 0.0297               |
|   | BayesNet   | 0.01                              | 379- (97.6804%)                       | 9 - (2.3196%)                           | 0.0322               |
|   | NaiveBayes | 0.01                              | 339- (87.3711%)                       | 49 - (12.6289 %)                        | 0.0894               |
|   | JRip       | 0.03                              | 377- (97.1649%)                       | 11 - (2.8351 %)                         | 0.0282               |
| <b>MATRIZ DE CONFUSION REPTree</b>  |            |                                   |                                       |   |                      |
| <pre> a   b   c   &lt;-- classified as 0   0   1       a = P 0   0   8       b = M 0   0 379       c = E </pre> |            |                                   |                                       |   |                      |

En la tabla 3.8 se puede observar que de todos los algoritmos que se aplicaron el (REPTree y BayesNet), nos dieron resultados idénticos siendo los valores más altos en las instancias clasificadas correctamente 379 – (97.6804%) y en cuanto a las instancias clasificadas incorrectamente 9 – (2.3196%) pero existe diferencia en el error absoluto medio siendo para REPTree igual a (0.0297) resultando menor que el que nos presenta BayesNet, debido a esto y a lo mencionado anteriormente acerca de la razón de precisión se determina, que el **REPTree** es el que nos da mejor precisión para el indicador del *nivel de participación en el curso*, en la misma tabla se indica la matriz de confusión del algoritmo con mayor precisión, donde las columnas indican las categorías clasificadas y las filas, las categorías reales de los datos teniendo a los elementos de la diagonal principal como los que ha acertado el clasificador y el resto son los errores.

*Tabla 3.9: Resultados obtenidos para el nivel de utilización de herramientas (informática)*

| INFORMATICA   |            |                                  |                                       |   |                      |  |
|---|------------|----------------------------------|---------------------------------------|---|----------------------|--|
| NIVEL DE UTILIZACION DE HERRAMIENTAS  | Algoritmo  | Tiempo en generar el modelo seg) | Instancias clasificadas correctamente | Instancias clasificadas incorrectamente | Error absoluto Medio |  |
|   | J48        | 0.01                             | 375- (96.6495%)                       | 13 - (3.3505%)                          | 0.0305               |  |
|   | REPTree    | 0.01                             | 378- (97.4227%)                       | 10 - (2.5773%)                          | 0.0309               |  |
|   | BayesNet   | 0.01                             | 367- (94.5876%)                       | 21 - (5.4124%)                          | 0.0407               |  |
|   | NaiveBayes | 0                                | 346- (85.1753%)                       | 42 - (10.8247%)                         | 0.0717               |  |
|   | JRip       | 0.02                             | 376- (96.6495%)                       | 12 - (3.3928%)                          | 0.0263               |  |
| <b>MATRIZ DE CONFUSION REPTree</b>  |            |                                  |                                       |   |                      |  |
| <pre> a   b   c   &lt;-- classified as 0   0   0       a = P 0   0   9       b = M 0   1-378     c = E </pre> |            |                                  |                                       |   |                      |  |

En la tabla 3.9 se tiene recopilados todos los resultados obtenidos por cada algoritmo referente al *indicador del nivel de la utilización de las herramientas*, aquí es notable que realizan una clasificación de las instancias con diferentes resultados ninguno coincide, se cuenta con un REPTree que clasifica las 378 instancias en un porcentaje del 97.4227%, pues es el que presenta la mejor precisión en la clasificación de instancias, además se tiene un NaiveBayes que es el que tiene un menor tiempo en generar el modelo, pero también es el que tiene un valor elevado en cuanto a instancias clasificadas erróneamente siendo éstas (42 - 10.8247%), en la tabla se presenta la matriz de confusión del algoritmo **REPTree** escogido como el idóneo para el indicador del nivel de utilización de las herramientas.

En vista de que ya se ha experimentado con cada uno de los algoritmos para los atributos y obteniendo los resultados mostrados en la tabla 3.8 y 3.9 conviene realizar la interpretación para poder seleccionar el que nos presente el algoritmo que nos ofrezca mejores resultados, ver la comparación de los resultados obtenidos en la tabla 3.10.

*Tabla 3.10: Comparación de la razón de precisión del algoritmo para cada indicador (informática)*

| Algoritmo  | Nivel de participación en el curso    |                      | Nivel de utilización de herramientas  |                      |
|------------|---------------------------------------|----------------------|---------------------------------------|----------------------|
|            | Instancias clasificadas correctamente | Error absoluto Medio | Instancias clasificadas correctamente | Error absoluto Medio |
| J48        | 377- (97.1649%)                       | 0.0311               | 375- (96.6495%)                       | 0.0305               |
| REPTree    | 379- (97.6804%)                       | 0.0297               | 378- (97.4227%)                       | 0.0309               |
| BayesNet   | 379- (97.6804%)                       | 0.0322               | 367- (94.5876%)                       | 0.0407               |
| NaiveBayes | 339- (87.3711%)                       | 0.0894               | 346- (85.1753%)                       | 0.0717               |
| JRip       | 377- (97.1649%)                       | 0.0282               | 376- (96.6495%)                       | 0.0263               |

En base a la comparación presentada se puede seleccionar el algoritmo que ofrece el mejor resultado para cada indicador, y en vista de que éstos fueron cada uno sujetos a análisis e interpretación, el siguiente paso a dar es dejar establecido cuáles son, en la siguiente tabla 3.11. Se mencionan:

*Tabla 3.11: Algoritmo seleccionado para cada indicador (informática)*

| Indicador                            | Algoritmo Seleccionado |
|--------------------------------------|------------------------|
| Nivel de participación en el curso   | REPTree                |
| Nivel de utilización de herramientas | REPTree                |

### 3.5.2 Resultados obtenidos para la carrera de *abogacía*

*Tabla 3.12: Resultados obtenidos para el nivel de participación en el curso (abogacía)*

| ABOGACIA  |            |                                   |                                       |   |                      |  |
|---|------------|-----------------------------------|---------------------------------------|---|----------------------|--|
| NIVEL DE PARTICIPACION EN EL CURSO  | Algoritmo  | Tiempo en generar el modelo (seg) | Instancias clasificadas correctamente | Instancias clasificadas incorrectamente | Error absoluto Medio |  |
|   | J48        | 0                                 | 395- (99.4962%)                       | 2 - (0.5038%)                           | 0.0067               |  |
|   | REPTree    | 0.01                              | 395- (99.4962%)                       | 2 - (0.5038%)                           | 0.0067               |  |
|   | BayesNet   | 0.01                              | 395- (99.4962%)                       | 2 - (0.5038%)                           | 0.0085               |  |
|   | NaiveBayes | 0                                 | 393- (98.9924%)                       | 4 - (10076%)                            | 0.0064               |  |
|   | JRip       | 0.03                              | 392- (98.7406%)                       | 5 - (1.2594%)                           | 0.0095               |  |
| MATRIZ DE CONFUSION J48   |            |                                   |                                       |   |                      |  |
| <pre> a   b   c   &lt;-- classified as 0   0   0     a = P 0   0   2     b = M 0   0 395     c = E </pre> |            |                                   |                                       |   |                      |  |

En la tabla 3.12 podemos observar que de todos los algoritmos que se aplicaron el (J48, REPTree y BayesNet), nos dieron resultados idénticos siendo los valores más altos en las instancias clasificadas correctamente 395 – (99.4962%) y en cuanto a las instancias clasificadas incorrectamente (2 – 0.5038%), pero existe diferencia en el error absoluto medio siendo para

J48 igual a (0.0297) resultando menor, al igual que en generar el modelo es el que tarda menos, que el que nos presenta REPTree y BayesNet, debido a esto y a la razón de precisión se determina, que el **J48** es el que nos da mejor precisión para el indicador del *nivel de participación en el curso*, en la misma tabla indicamos la matriz de confusión del algoritmo con mayor precisión, donde las columnas indican las categorías clasificadas y las filas, las categorías reales de los datos teniendo a los elementos de la diagonal principal como los que ha acertado el clasificador y el resto son los errores.

*Tabla 3.13: Resultados obtenidos para el nivel de utilización de herramientas (abogacía)*

| ABOGACIA   |            |                                   |                                       |   |                      |  |
|--|------------|-----------------------------------|---------------------------------------|---|----------------------|--|
| NIVEL DE UTILIZACION DE HERRAMIENTAS   | Algoritmo  | Tiempo en generar el modelo (seg) | Instancias clasificadas correctamente | Instancias clasificadas incorrectamente | Error absoluto Medio |  |
|  | J48        | 0                                 | 395- (99.4962%)                       | 2 - (0.5038%)                           | 0.0067               |  |
|  | REPTree    | 0.01                              | 395- (99.4962%)                       | 2 - (0.5038%)                           | 0.0067               |  |
|  | BayesNet   | 0.01                              | 395- (99.4962%)                       | 2 - (0.5038%)                           | 0.0085               |  |
|  | NaiveBayes | 0                                 | 391- (98.4887%)                       | 6 - (1.5113%)                           | 0.009                |  |
|  | JRip       | 0                                 | 393- (98.9924%)                       | 4 - (1.0076%)                           | 0.0067               |  |
| <b>MATRIZ DE CONFUSION J48</b>   |            |                                   |                                       |   |                      |  |
| <pre> a  b  c  &lt;-- classified as 0  0  0     a = P 0  0  2     b = M 0  0 395    c = E </pre> |            |                                   |                                       |   |                      |  |

En la tabla 3.13 se tiene recopilados todos los resultados obtenidos por cada algoritmo referente al *indicador del nivel de la utilización de las herramientas*, aquí es notable que realizan una clasificación de las instancias con resultados similares en el caso de los algoritmos (J48, REPTree y BayesNet), presentan para el caso de las instancias clasificadas correctamente 395- (99.4962%), mientras que para las instancias clasificadas incorrectamente 2 - (0.5038%), pero se diferencian en el tiempo en generar el modelo pues el algoritmo J48 con un error absoluto medio de (0.0067), es el que presenta el menor tiempo en comparación al algoritmo REPTree y BayesNet , además de ser el que indica un error absoluto medio más bajo de entre

los algoritmos (BayesNet, NaiveBayes y JRip), en la tabla se presenta la matriz de confusión del algoritmo **J48** escogido como el idóneo para el identificador del nivel de utilización de las herramientas.

Ya que se ha experimentado con cada uno de los algoritmos para los atributos y obteniendo los resultados mostrados en la tabla 3.12 y 3.13 queda realizar la interpretación para poder seleccionar el que nos presente el algoritmo que nos ofrezca mejores resultados, ver la comparación de los resultados obtenidos en la tabla 3.14

*Tabla 3.14: Comparación de la razón de precisión del algoritmo para cada indicador (abogacía)*

|                   | Nivel de participación en el curso    |                      | Nivel de utilización de herramientas  |                      |
|-------------------|---------------------------------------|----------------------|---------------------------------------|----------------------|
| Algoritmo         | Instancias clasificadas correctamente | Error absoluto Medio | Instancias clasificadas correctamente | Error absoluto Medio |
| <b>J48</b>        | 395- (99.4962%)                       | 0.0067               | 395- (99.4962%)                       | 0.0067               |
| <b>REPTree</b>    | 395- (99.4962%)                       | 0.0067               | 395- (99.4962%)                       | 0.0067               |
| <b>BayesNet</b>   | 395- (99.4962%)                       | 0.0085               | 395- (99.4962%)                       | 0.0085               |
| <b>NaiveBayes</b> | 393- (98.9924%)                       | 0.0064               | 391- (98.4887%)                       | 0.009                |
| <b>JRip</b>       | 392- (98.7406%)                       | 0.0095               | 393- (98.9924%)                       | 0.0067               |

En base a la comparación presentada en la tabla 3.14 se puede seleccionar el algoritmo que ofrece el mejor resultado para cada indicador, y en vista de que éstos fueron cada uno sujetos a análisis e interpretación, el siguiente paso a dar es dejar establecido cuáles son, en la siguiente tabla 3.15. Se mencionan:

Tabla 3.15: Algoritmo seleccionado para cada indicador(abogacía)

| Indicador                            | Algoritmo Seleccionado |
|--------------------------------------|------------------------|
| Nivel de participación en el curso   | J48                    |
| Nivel de utilización de herramientas | J48                    |

### 3.5.3 Comparación de las matrices de confusión entre la carrera de *informática* y *abogacía*

En la tabla 3.16 se plasman las matrices de confusión para cada carrera con el objetivo de realizar un análisis comparativo de los resultados que proyectan en cada uno de los indicadores *nivel de participación en el curso* y *nivel de utilización de las herramientas*.

Tabla 3.16: Comparación de las matrices de confusión

| Nivel de participación en el curso |   |          |                   | Nivel de utilización de las herramientas |   |          |                   |   |   |     |                   |   |   |     |                   |
|------------------------------------|---|----------|-------------------|--|---|----------|-------------------|---|---|-----|-------------------|---|---|-----|-------------------|
| Informática                        |   | Abogacía |                   | Informática                              |   | Abogacía |                   |   |   |     |                   |   |   |     |                   |
| a                                  | b | c        | <-- classified as | a  | b | c        | <-- classified as | a | b | c   | <-- classified as | a | b | c   | <-- classified as |
| 0                                  | 0 | 1        | a = P             | 0  | 0 | 0        | a = P             | 0 | 0 | 0   | a = P             | 0 | 0 | 0   | a = P             |
| 0                                  | 0 | 8        | b = M             | 0  | 0 | 2        | b = M             | 0 | 0 | 9   | b = M             | 0 | 0 | 2   | b = M             |
| 0                                  | 0 | 379      | c = E             | 0  | 0 | 395      | c = E             | 0 | 1 | 378 | c = E             | 0 | 0 | 395 | c = E             |

- ❖ **Nivel de participación en el curso:** En la carrera de *informática* se nota que de los 379 estudiantes que presentan un nivel de participación en el curso Escaso (E), todos son correctamente seleccionados, en el caso del nivel Moderado (M) presenta la cantidad de 8 estudiantes que son, los que tienen un nivel medio de participación o interés en el curso, pues resulta ser que en informática existen más estudiantes que participan en el curso en un *nivel Moderado*, que los de la carrera de abogacía ya que aquí son apenas 2 los que presentan esta incidencia y para el nivel Permanente (P), en informática se tiene la representación de 1 solo estudiante, mientras que para la carrera de abogacía no

existe representación alguna observando que la cantidad de estudiantes que acogen este nivel es igual a 0, en base a todas estas observaciones, se puede concluir en general que el nivel predominante en los estudiantes es el **Escaso**, tanto para los de la carrera de informática como para los de abogacía, ya que en ambos casos la gran mayoría de estudiantes se encuentran concentrados en ese nivel.

- ❖ **Nivel de utilización de las herramientas:** Aquí se nota que en la carrera de *informática* de los 379 estudiantes que presentan un nivel de participación en el curso Escaso (E), 378 son clasificados correctamente y 1 con error, en lo que concierne al nivel Moderado (M) se aprecia que existen 9 estudiantes que hacen uso de las herramientas en un nivel medio, caso contrario al de los estudiantes de la carrera de *abogacía* ya que solamente 2 estudiantes presentan actitudes en éste nivel mientras que para el nivel Permanente (P), ambas carreras se encuentran en iguales condiciones ya que tienen una incidencia de estudiantes igual a 0. En base a lo expuesto, se puede concluir de manera general que los estudiantes de la carrera de informática como de abogacía presentan mayor incidencia en el nivel de utilización de las herramientas **Escaso**.

## 3.6 Experimentación con algoritmo de clustering

Ahora experimentaremos con el algoritmo caracterizado por su sencillez *SimpleKMeans* además de ser el más utilizado para hacer clustering y al extenso material bibliográfico con el que se cuenta sobre este algoritmo, teniendo en cuenta que es necesario realizar varios experimentos con este para obtener el resultado más óptimo posible, hacemos uso de éste ya que nos interesa evaluar cada atributo que conforma el *nivel de utilización de herramientas* para lo cual se ha elegido dos grupos de población, que hacen referencia a dos asignaturas en específico *Lógica de la programación* (informática) y *Ética y derechos humanos* (abogacía) a estas asignaturas se les aplicará el algoritmo, utilizando los valores reales obtenidos para cada atributo que conforman este indicador del nivel de utilización de las herramientas de la plataforma conforme las acciones que éstos realizan sobre las mismas, de tal manera para posteriormente poder proceder a realizar las comparaciones pertinentes entre estas dos asignaturas de la carrera de Informática y abogacía.

### 3.6.1 Experimentación y resultados del curso de lógica de la programación

Para el **primer experimento**:

- Desde la aplicación WEKA, seleccionamos el algoritmo clúster SimpleKmeans y le configuramos el valor de 3 clústeres ya que se cuenta con un número considerable de instancias *Lógica de la programación* (205), para obtener tres grupos definidos de estudiantes de acuerdo a las acciones que éstos realizan, teniéndolos en la siguiente clases de actividad en la utilización de las herramientas que presenta la plataforma y estas son Escasa(E) representado por los valores mínimos, Moderado(M) los valores que se encuentran en un rango medio y Permanente(P) los valores en un rango alto. Dejando los demás valores de configuración por defecto, así como el modo de evaluación del clúster le dejamos en *Use training set* (Morate, D.) en esta opción WEKA

entrenara el método con todos los datos disponibles y luego lo aplicará otra vez sobre los mismos. En la tabla 3.17 podemos observar el número de iteraciones que se han efectuado para obtener los clústeres, también tenemos la suma de los errores cuadrados, también las instancias que corresponden a cada clúster.

*Tabla 3.17: Primer experimento con el algoritmo SimpleKMeans – semilla 1000 (lógica de la programación)*

| Lógica de la programación |   |                         |  |  |
|---------------------------|---|-------------------------|--|--|
| Algoritmo                 | Clúster   | Instancias clasificadas |  |  |
| SimpleKMeans              | 0   | 57 – (28%)              |  |  |
|                           | 1   | 16 – (8%)               |  |  |
|                           | 2   | 132 – (64%)             |  |  |
|                           | <pre> KMeans *****  Number of iterations: 11 Within cluster sum of squared errors: 69.0410193339927 Missing values globally replaced with mean/mode  Cluster centroids: Attribute          Full Data      Cluster#                    (205)         0             1             2                    (57)         (16)         (132) ----- view_forum_forum   15.6341        36.5789       11.5         7.0909 add_post_forum     1.6146         3.9825        1.125        0.6515 view_discussion_forum 16.9317       43.2982       11.0625      6.2576 update_postforum   0.3171         0.8246        0.4375      0.0833 add_discussion     0.8829         1.9649        0.625        0.447 view_resource      3.7854         6.7018        4.875        2.3939 view_assignment    7.3317         13.1579       11.0625     4.3636 upload_upload      2.3171         4.4386        2.5          1.3788 ms_twt             2.5415         6.0702        1.875        1.0985 write_message      3.7561         9.2281        3.375        1.4394 history_message    4.2732         9.0702        2.0625      2.4697 view_quiz          9.5854         16.6667       10.6875     6.3939 num_intevaladpb    0.9902         1.193         1.0625      0.8939 tie_evaladpb       8487.1659     2392.7895     85120.125   1829.9848 num_intevaladsb    0.6341         0.8947        0.8125      0.5 tie_evaladsb       2726.722     4503.5439     6974.6875   1444.553 num_accesos_curso 53.922        115.2807      50.8125     27.803  Clustered Instances  0      57 ( 28%) 1      16 (  8%) 2     132 ( 64%) </pre> |                         |  |  |

Además se puede apreciar los resultados alcanzados del primer experimento con el algoritmo SimpleKMeans, en el clúster 1 se notó que difiere con carencia de más de la mitad en instancias obtenidas en relación al 0 y 2 ya que en éstos también hay una diferencia puesto que el clúster 0 tiene el 28% de instancias y el clúster 2 del 64% mientras que el clúster 1 si es apenas del 8% lo que equivale a 16 instancias, también en la suma de los errores cuadrados nos proyecta un valor de 69.04, pues es muy elevado y se debe a que se cuenta con un número de atributos considerable porque se probó con menos atributos y la suma de los errores cuadrados disminuyo, pero necesitamos todos los atributos que se muestran en el experimento porque éstos son los implicados en el nivel de utilización de las herramientas, además en este experimento no se manifiesta una agrupación, con características similares, que es lo que se busca en base a los atributos descubrir y determinar los grupos que **presenten similitud en la incidencia de alguna acción que realicen sobre las herramientas que dispone el eva**, por esta razón conviene realizar otro experimento con el algoritmo SimpleKMeans.

Para el **segundo experimento**:

- Se mantiene el mismo número de clúster ya que desde un inicio se definió tres clases de actividad que según los resultados se ubicarán en la clase que correspondan, lo que se configura y difiere del anterior experimento es el aumento del valor de seed (semilla) para dar valores a los centros iniciales, pasaremos de 1000 a aplicar un valor de 3000, pues con el objetivo de encontrar aquella que minimice la suma del error cuadrático, aunque no garantiza una semilla óptima pero si una buena asignación de instancias que es para lo que precisamente sirve este método heurístico<sup>10</sup>.

---

<sup>10</sup> Heurístico. Duhalde y González (1997) señalan que es “un procedimiento que ofrece la posibilidad de seleccionar estrategias que nos acercan a una solución” <http://bit.ly/ri3LpO>

En la tabla 3.18 se pueden apreciar los resultados de un segundo experimento.

*Tabla 3.18: Segundo experimento con el algoritmo SimpleKMeans – semilla 3000 (lógica de la programación)*

| Lógica de la programación |   |                         |
|---------------------------|---|-------------------------|
| Algoritmo                 | Clúster   | Instancias clasificadas |
| SimpleKMeans              | 0   | 70 - (34%)              |
|                           | 1   | 119 - ( 58%)            |
|                           | 2   | 16 - (8%)               |
|                           | <pre> kMeans *****  Number of iterations: 5 Within cluster sum of squared errors: 69.45679640189276 Missing values globally replaced with mean/mode  Cluster centroids:  Attribute          Full Data      Cluster#                    (205)         0           1           2                    (70)        (119)       (16) ----- view_forum_forum   15.6341      33.1286     5.8992     11.5 add_post_forum     1.6146       3.5857      0.521      1.125 view_discussion_forum 16.9317     38.9286     4.7815     11.0625 update_postforum   0.3171       0.7286     0.0588     0.4375 add_discussion     0.8829       1.8571     0.3445     0.625 view_resource      3.7854       6.1286     2.2605     4.875 view_assignment    7.3317      12.7286     3.6555     11.0625 upload_upload      2.3171       4.3286     1.1092     2.5 ms_twt             2.5415       5.4143     0.9412     1.875 write_message      3.7561       8.0714     1.2689     3.375 history_message    4.2732       7.8857     2.4454     2.0625 view_quiz          9.5854       15.4        6.0168     10.6875 num_invaladpb     0.9902       1.1857     0.8655     1.0625 tie_evaladpb     8487.1659    2135.9714   1919.5714   85120.125 num_invaladab     0.6341       0.9429     0.4286     0.8125 tie_evaladab     2726.722     3822.4857   1511        6974.6875 num_accesos_curso 53.922      104.2429    24.7395     50.8125  Clustered Instances  0      70 ( 34%) 1     119 ( 58%) 2      16 (  8%) </pre> |                         |

Aquí apreciamos una agrupación en la que el clúster 1 es más alta que el clúster 0 y 2 mientras que el clúster 2 tiene un porcentaje de instancias menor del 8% en comparación al clúster 0 y 1, además se nota un aumento en la suma del error

cuadrático a 69.45 aunque no es una diferencia notable pues es con decimas en relación al primer experimento esta distribución de las instancias no convence debido a que dificulta la interpretación de los resultados, lo que convierte a esta agrupación como no indicada en cuanto a la optima definición de los grupos, lo que obliga a seguir realizando un intento más de experimento, que por lo menos se acerque a obtener clúster ubicados en cada rango de valores, mínimos, medios y altos.

Para el **tercer experimento**:

Esta vez cabe indicar que se realizó algunas pruebas en cuanto a variar en el valor de la semilla, es por ello que la que se ha optado por escoger el valor de 7000, los demás valores continúan igual que en los experimentos anteriores.

En la tabla 3.19, se detallan los resultados que proporcionó el algoritmo y se muestran a continuación.

**Tabla 3.19: Tercer experimento con el algoritmo SimpleKMeans – semilla 7000 (lógica de la programación)**

| Lógica de la programación |   |                         |  |  |
|---------------------------|---|-------------------------|--|--|
| Algoritmo                 | Clúster   | Instancias clasificadas |  |  |
| SimpleKMeans              | 0   | 36 - (18%)              |  |  |
|                           | 1   | 83 - (40%)              |  |  |
|                           | 2   | 86 - (42%)              |  |  |
|                           | <pre> Number of iterations: 5 Within cluster sum of squared errors: 69.31334547797475 Missing values globally replaced with mean/mode  Cluster centroids:  Attribute          Full Data      Cluster#                    (205)         0           1           2                    (205)         (36)        (83)        (86) ----- view_forum_forum   15.6341        42.0833     17.241     3.0116 add_post_forum     1.6146         4.8889     1.7108     0.1512 view_discussion_forum 16.9317       53.8333     15.9518     2.4302 update_postforum   0.3171         1.1389     0.2771     0.0116 add_discussion     0.8829         1.8611     1.2771     0.093 view_resource      3.7854         7.3889     4.4096     1.6744 view_assignment    7.3317         15.5278     9.4578     1.8488 upload_upload      2.3171         4.5833     3.3494     0.3721 ms_tweet           2.5415         7.6944     2.2771     0.6395 write_message      3.7561         11.4722     3.3494     0.9186 history_message    4.2732         11         3.2892     2.407 view_quiz          9.5854         17.3333     11.8916     4.1163 num_intevaladpb   0.9902         1.1667     1.1446     0.7674 tie_evaladpb      9348.0419      2624.2963  12767.7756  8862.1924 num_intevaladb    0.6341         0.9167     0.9036     0.2558 tie_evaladb       4729.2292      7185.0938  4420.2588  3999.3852 num_accesos_curso 53.922         135.5833   57.4217    16.3605  Clustered Instances  0      36 ( 18%) 1      83 ( 40%) 2      86 ( 42%) </pre> |                         |  |  |

Como ya se mencionó anteriormente pues se ha seleccionado ese valor en la semilla puesto que es el que presenta más semejanza entre las distribuciones de instancias para cada clase, en el clúster 0 se tiene las instancias con valores relativamente mayores al clúster 1 y 2, lo que resulta ideal para la interpretación ya que no contamos con ambigüedades como en las agrupaciones de los experimentos anteriores que existían,

algunos atributos mayores en el clúster 1 y otros mayores en el clúster 0. El caso particular del centroide que pertenece al atributo *view\_quiz* se encuentra un valor de 17.3333, este es el valor mayor entre el clúster 1 y 2 es decir en este clúster se agrupan los estudiantes que **mayor cantidad de veces visualizan la herramienta de cuestionario**, se habla de este en particular debido a que los estudiantes acceden o ingresan a cuestionario no necesariamente para rendir una evaluación ya que después de haber rendido el examen también pueden volver a ingresar para revisar en que pregunta fallaron, es por esto que el número de visualización de los cuestionarios es mucho mayor que el atributo del número de intentos por resolver las evaluaciones a distancia tanto para el primer bimestre como para el segundo (*num\_intevaladspb*, *num\_intevaladsb*), aunque a excepción del atributo *tie\_evaladpb* (tiempo que tarda en rendir la evaluación del primer bimestre) que tiene el valor más alto en el clúster 1, pero el *num\_intevaladspb* (número de intentos en la evaluación a distancia del primer bimestre), con un valor medio en base a esto se puede concluir que los estudiantes con un valor medio en el número de intentos por resolver la evaluación a distancia del primer bimestre, utilizan mayor cantidad del tiempo, pues denota que en vez de realizar varios intentos por dar respuesta, **a las preguntas se toman el tiempo de analizar y leer más detenidamente las preguntas para proceder a dar contestación**. En los atributos que restan los que están agrupados en el clúster 0 poseen los valores más altos y más adelante se realiza una interpretación del nivel de utilización de cada herramienta.

En base a los resultados obtenidos, tenemos que el clúster 0 es el que posee los valores más altos en sus atributos, encontrándose aquí los estudiantes que tienen un nivel de utilización de las herramientas *permanente*, mientras que el clúster 1 es el que tiene los valores medios pues esto conlleva a situar a los estudiantes que tienen un nivel de utilización de las herramientas *moderado*, y en cuanto al clúster 2 aquí están los

valores bajos a lo que corresponden los estudiantes con un nivel de utilización de las herramientas *escaso*.

A continuación se evalúan los resultados de los tres experimentos realizados para de esta manera establecer cuál de éstos es el que da mejores resultados.

*Tabla 3.20: Comparativa de los experimentos realizados utilizando SimpleKMeans (lógica de la programación)*

| SimpleKMeans               | Clúster | Instancias Clasificadas | Suma de errores cuadrados |
|----------------------------|---------|-------------------------|---------------------------|
| <b>Primer Experimento</b>  | 0       | 57 - (28%)              | 69.04                     |
|                            | 1       | 16 - (8%)               |                           |
|                            | 2       | 132 - (64%)             |                           |
| <b>Segundo Experimento</b> | 0       | 70 - (34%)              | 69.45                     |
|                            | 1       | 119 - (58%)             |                           |
|                            | 2       | 16 - (8%)               |                           |
| <b>Tercer Experimento</b>  | 0       | 36 - (18%)              | 69.31                     |
|                            | 1       | 83 - (40%)              |                           |
|                            | 2       | 86 - (42%)              |                           |

En la tabla 3.20 se aprecia la recopilación de todos los resultados de cada experimento, encontrando que los valores obtenidos del primer experimento presenta resultados aceptables entre el clúster 0 y 2 pero en la suma de errores cuadrados es el que tiene el valor menor de todos los experimentos pero su distribución de instancias es confusa por lo que no resulta conveniente para ser sujeto a análisis e interpretación, en el segundo experimento no se encuentra una agrupación semejante entre los clúster por esto también fue desechado y se optó por seguir probando; llegando así a un tercer experimento aquí ya se contó con una mejor distribución de las instancias en la agrupación además de que posee el menor valor en la suma

de errores cuadrados, respecto al experimento dos reflejando ser el más adecuado para hacer la interpretación acorde con los resultados basados en los grupos.

- ❖ **Utilización de herramientas**, para este indicador se consideran los siguientes atributos en cada clúster, aquí se muestra el análisis para obtener el nivel de utilización de las herramientas por parte de los estudiantes:

Tabla 3.21: Nivel de utilización de las herramientas (lógica de la programación)

| Herramienta | Clúster | Atributo                     | Centroide | Nivel utilización |
|-------------|---------|------------------------------|-----------|-------------------|
| Foros       | 0       | <i>view_forum_forum</i>      | 42.0833   | Permanente        |
|             |         | <i>add_post_forum</i>        | 4.8889    |                   |
|             |         | <i>view_discussion_forum</i> | 53.8333   |                   |
|             |         | <i>update_postforum</i>      | 1.1389    |                   |
|             |         | <i>add_discussion</i>        | 1.8611    |                   |
|             | 1       | <i>view_forum_forum</i>      | 17.241    | Moderado          |
|             |         | <i>add_post_forum</i>        | 1.7108    |                   |
|             |         | <i>view_discussion_forum</i> | 15.9518   |                   |
|             |         | <i>update_postforum</i>      | 0.2771    |                   |
|             |         | <i>add_discussion</i>        | 1.2771    |                   |
|             | 2       | <i>view_forum_forum</i>      | 3.0116    | Escaso            |
|             |         | <i>add_post_forum</i>        | 0.1512    |                   |
|             |         | <i>view_discussion_forum</i> | 2.4302    |                   |
|             |         | <i>update_postforum</i>      | 0.0116    |                   |
|             |         | <i>add_discussion</i>        | 0.093     |                   |
| Recursos    | 0       | <i>view_resource</i>         | 7.3889    | Permanente        |
|             | 1       | <i>view_resource</i>         | 4.4096    | Moderado          |
|             | 2       | <i>view_resource</i>         | 1.6744    | Escaso            |
| Tareas      | 0       | <i>view_assignment</i>       | 15.5278   | Permanente        |
|             |         | <i>upload_upload</i>         | 4.5833    |                   |
|             | 1       | <i>view_assignment</i>       | 9.4578    | Moderado          |
|             |         | <i>upload_upload</i>         | 3.3494    |                   |
|             | 2       | <i>view_assignment</i>       | 1.8488    | Escaso            |
|             |         | <i>upload_upload</i>         | 0.3721    |                   |
|             | 0       | <i>write_message</i>         | 11.4722   | Permanente        |
|             |         | <i>history_message</i>       | 11        |                   |
|             | 1       | <i>write_message</i>         | 3.3494    | Moderado          |

|                     |   |                        |            |            |
|---------------------|---|------------------------|------------|------------|
| <b>Mensajería</b>   |   | <i>history_message</i> | 3.2892     |            |
|                     | 2 | <i>write_message</i>   | 0.9186     | Escaso     |
|                     |   | <i>history_message</i> | 2.407      |            |
| <b>Twitter</b>      | 0 | <i>ms_twitter</i>      | 5.6721     | Permanente |
|                     | 1 | <i>ms_twitter</i>      | 1.5576     | Moderado   |
|                     | 2 | <i>ms_twitter</i>      | 0.642      | Escaso     |
| <b>Cuestionario</b> | 0 | <i>num_intevaldpb</i>  | 1.1667     | Permanente |
|                     |   | <i>tie_evaladpb</i>    | 2624.2963  |            |
|                     |   | <i>num_intevaldsb</i>  | 0.9167     |            |
|                     |   | <i>tie_evaladsb</i>    | 7185.0938  |            |
|                     |   | <i>view_quiz</i>       | 17.3333    |            |
|                     | 1 | <i>num_intevaldpb</i>  | 1.1446     | Moderado   |
|                     |   | <i>tie_evaladpb</i>    | 12767.7756 |            |
|                     |   | <i>num_intevaldsb</i>  | 0.9036     |            |
|                     |   | <i>tie_evaladsb</i>    | 4420.2588  |            |
|                     |   | <i>view_quiz</i>       | 11.8916    |            |
|                     | 2 | <i>num_intevaldpb</i>  | 0.7674     | Escaso     |
|                     |   | <i>tie_evaladpb</i>    | 8862.1924  |            |
|                     |   | <i>num_intevaldsb</i>  | 0.2558     |            |
|                     |   | <i>tie_evaladsb</i>    | 3999.3852  |            |
|                     |   | <i>view_quiz</i>       | 4.1163     |            |

En la tabla 3.21 se presenta los niveles de utilización de las herramientas por parte de los estudiantes, se realizó la asignación de los atributos con su correspondiente herramienta en cada clúster.

Además se puede apreciar, en base a los resultados que ahí se exhiben contamos con tres grupos, grupo 1 (clúster 0) corresponde a los estudiantes que tienen un nivel de utilización de la herramienta foros, recursos, tareas, mensajería, twitter y cuestionario *permanente*, grupo 2 (clúster 1) está conformado por los estudiantes que tienen un nivel de utilización de la herramienta foros, recursos, tareas, mensajería, twitter y cuestionario *moderado*, por último tenemos el grupo 3 (clúster 2) aquí pertenecen los estudiantes que poseen un nivel de utilización de la herramienta foros, recursos, tareas, mensajería, twitter y cuestionario *escaso*.

A continuación se presentan las gráficas obtenidas con el algoritmo *SimpleKMeans* además se dará una interpretación de estas.

### FOROS

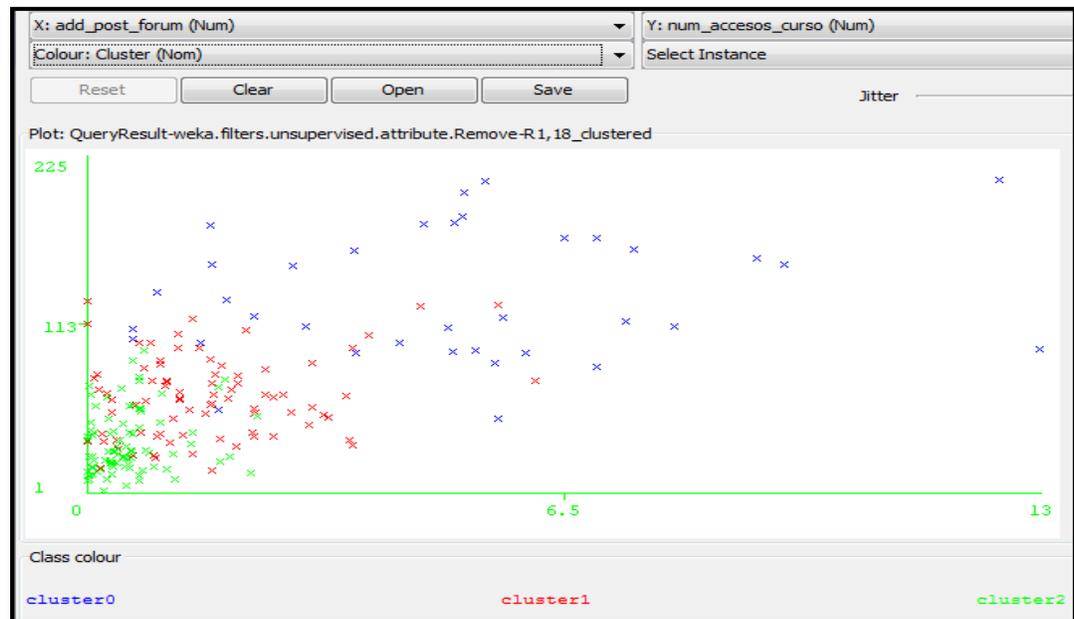


Figura 3.12: Número de veces que contesta un foro según el número de accesos al curso (lógica de la programación)

En la figura 3.12 el número de veces que contesta un foro se encuentra en el eje  $x$  éstos están representados en los clúster con tres colores diferentes para el clúster 0 azul, para el clúster 1 rojo, y para el clúster 2 verde, aquí podemos apreciar que la mayoría de los estudiantes presentan un patrón que mientras **más acceden al curso, mayor es el número de veces que contestan un foro** observando en el eje de las  $x$  grupo 1 (clúster 0) se puede apreciar esta tendencia de los estudiantes lo que indica que poseen un *nivel de utilización de la herramienta* alto (*permanente*), grupo 2 (clúster 1) comprende los estudiantes que tienen un *nivel de utilización de la herramienta* medio (*moderado*), y los que conservan un *nivel de utilización de la herramienta* bajo (*escaso*) están en el grupo 3 (clúster 2).

## RECURSOS

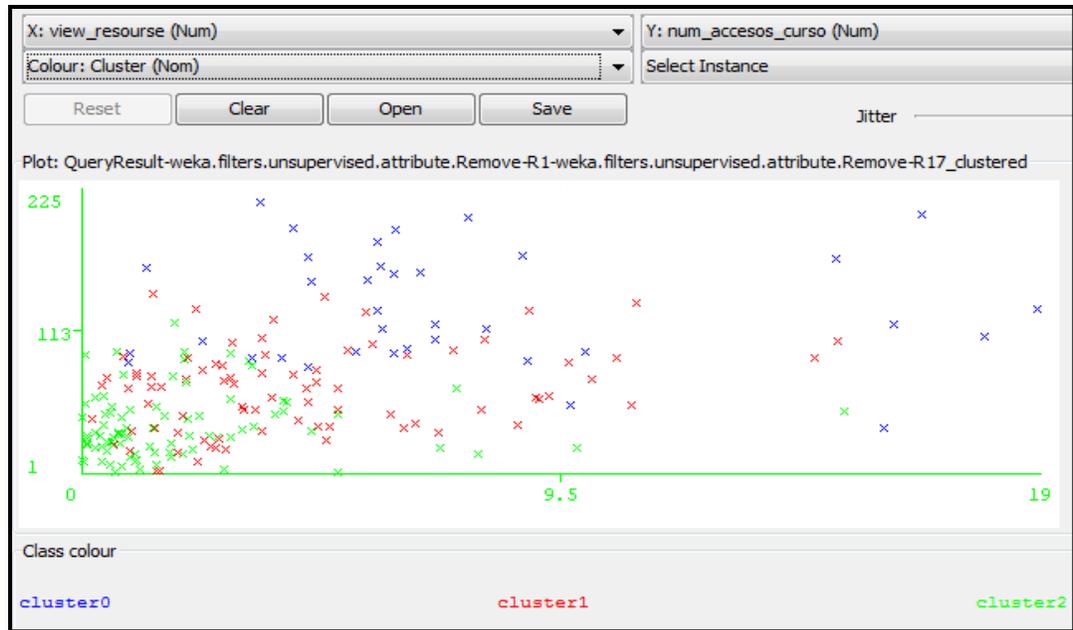


Figura 3.13: Número de veces que descarga un recurso según el número de accesos al curso (lógica de la programación)

En la figura 3.13 se aprecia la interrelación entre los atributos *view\_resource* (número de veces que descarga un recurso) en el eje de la x, según el número de accesos al curso en el eje de la y.

En la gráfica es evidente que el grupo 1 (clúster 0), es el que posee los valores mayores pues se nota que los estudiantes tienen mayor tendencia para **descargar un recurso, cuando acceden o visitan más el curso**, los demás grupos pertenecientes al clúster 1 están en un nivel medio y en cuanto al clúster 2 sus resultados son bastante inferiores.

## TAREAS

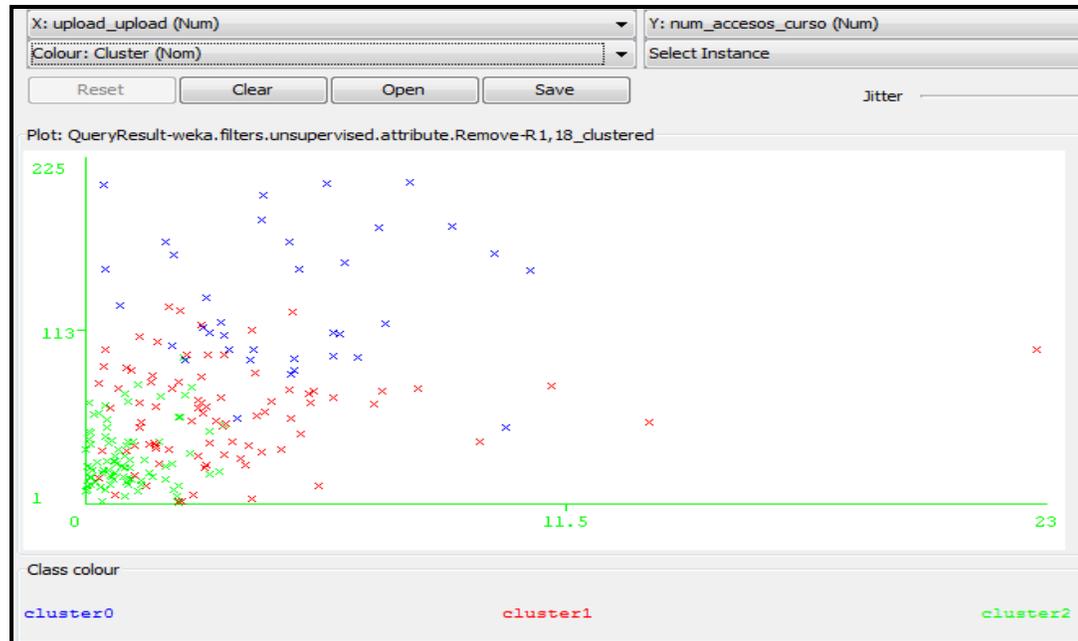


Figura 3.14: Número de veces que envía o sube una tarea según el número de accesos al curso (lógica de la programación)

En la figura 3.14 se aprecia los atributos *upload\_upload* (número de veces que sube una tarea) eje de la x, número de accesos al curso eje de la y.

En la gráfica se nota que los estudiantes del grupo 1 (clúster 0), tienden a **enviar o subir mayor número de tareas, cuando ingresan o acceden mayor cantidad de veces al curso**, para enterarse de cuáles son las que tienen que enviar, es notorio que los valores más altos están en el grupo 1, lo que permite determinar que se encuentran ubicados aquí los estudiantes que presentan un *nivel de utilización de la herramienta* permanente, le sigue el grupo 2 (clúster 1), que de entre los tres, se observa que, es el que tiene los valores medios y por último está el grupo 3 (clúster 2), que refleja los valores menos notorios en cuanto a la utilización de la herramienta.

## MENSAJERIA

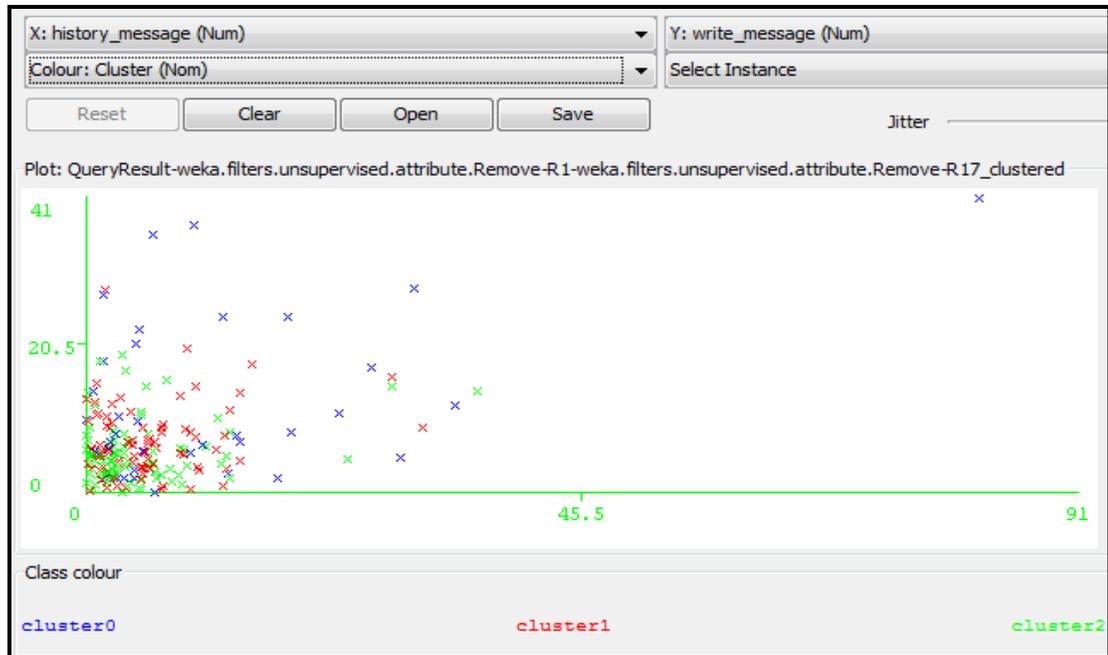
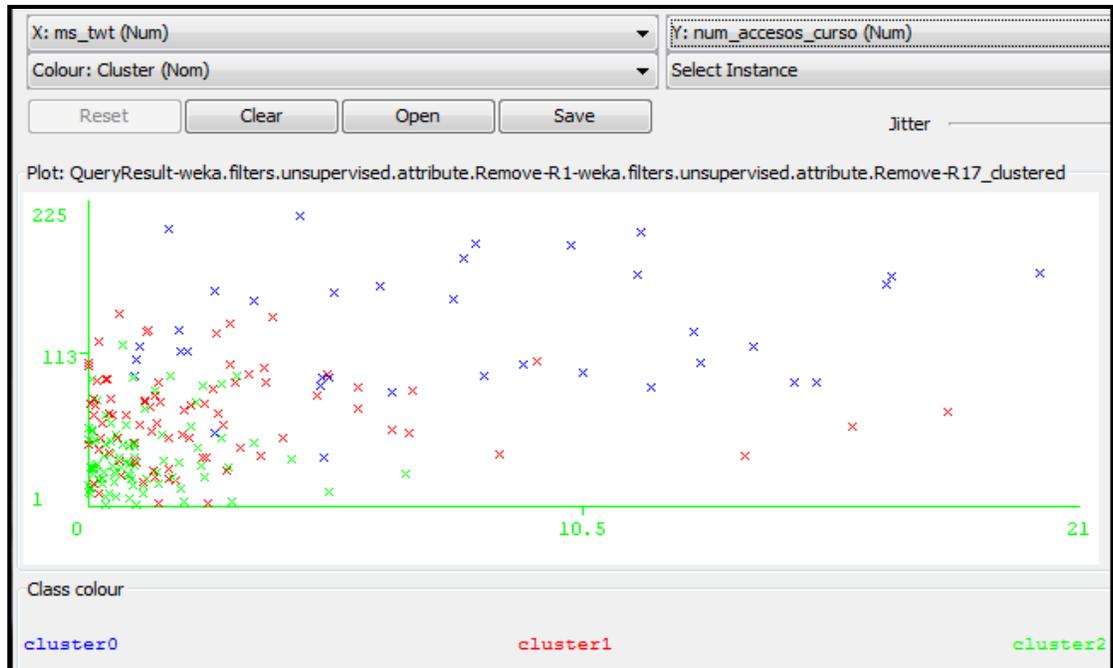


Figura 3.15: Número de veces que revisa el historial de mensajes según número de mensajes que ha escrito (lógica de la programación)

En la figura 3.15 se tiene los atributos *history\_message* (número de veces que revisa el historial de mensajes), y en el eje de la *y* *write\_message* (número de mensajes que ha escrito).

En la gráfica se puede apreciar que los estudiantes tienden a **revisar más historial de mensajes, mientras sea mayor el número de mensajes** que han escrito, puesto que resulta lógico, ya que si un estudiante envía un mensaje pues esta a la espera de alguna respuesta lo que obliga a estar continuamente revisando directamente al historial de mensajes para leer las posibles contestaciones, pues se nota que los valores más altos los posee el grupo 1 (clúster 0), encajándolo en un *nivel de utilización de la herramienta* permanente.

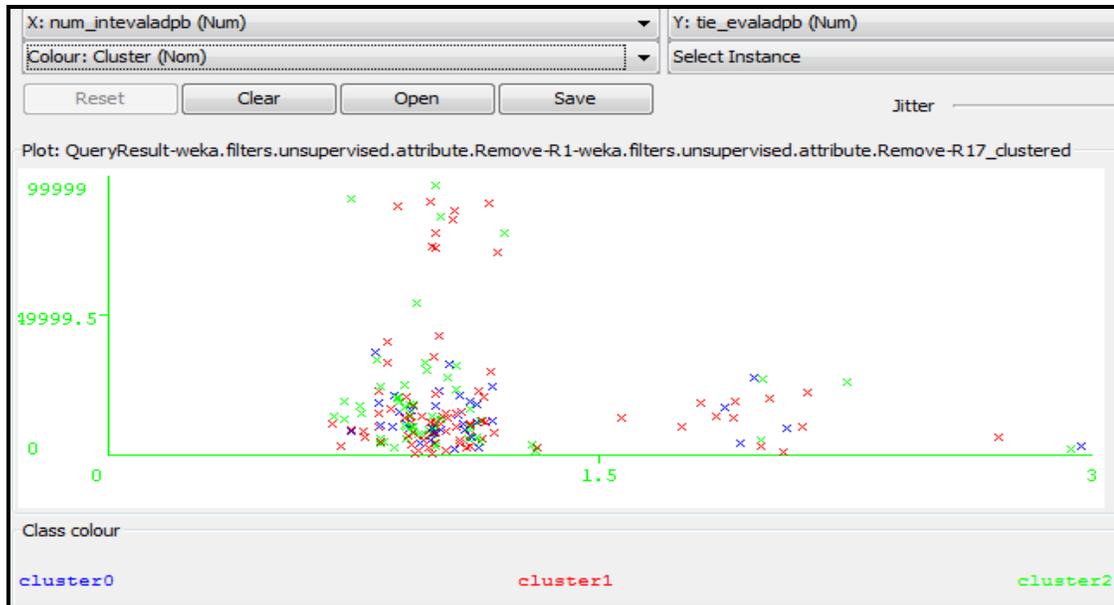
## TWITTER



*Figura 3.16: Número de mensajes twitter según el número de accesos al curso (lógica de la programación)*

En la figura 3.16 se aprecia en el eje de la y el número de accesos al curso, con el eje de la x *ms\_twt* (número de mensajes twitter). Se observa que existe una incidencia en cada grupo siendo así que para el grupo 1 (clúster 0) se observa que mientras **los valores son altos en el número de accesos al curso, mayor es el número de mensajes twitter enviados** por parte de los estudiantes, lo que permite establecer que a este grupo lo comprenden aquellos estudiantes que tienen un *nivel de utilización de la herramienta* permanente en contraposición a este se encuentra el grupo 2 (clúster 1) que para la mayoría de las instancias posee un valor medio, y en el grupo 3 (clúster 2) se considera que el número de accesos al curso es bajo, al igual que el número de mensajes twitter.

## CUESTIONARIO



*Figura 3.17: Número de intentos por rendir una evaluación según el tiempo utilizado (evaluación a distancia primer bimestre), (lógica de la programación)*

En la figura 3.17 se observan los atributos *num\_intevaladpb* (número de intentos en la evaluación a distancia del primer bimestre) eje de la x, y en lo que corresponde al atributo *tie\_evaladpb* (tiempo que tarda en la evaluación a distancia del primer bimestre) en el eje de la y.

En la gráfica se aprecia una comparación del tiempo utilizado para rendir la evaluación a distancia del primer bimestre con el número de intentos por resolver la evaluación, encontrando así algunas tendencias en los estudiantes con respecto a la herramienta cuestionario, en la gráfica para el grupo 1 (clúster 0) se nota una instancia que posee un valor medio para el tiempo utilizado, pero el número de intentos es mayor, grupo 2 (clúster 1) se observan varias instancias y son las que tienen los valores altos en el tiempo utilizado para rendir la evaluación a distancia del primer bimestre, mientras que para el número de intentos por resolver la evaluación existen unas que tienen un valor alto y otras medio. Por lo que indica para este grupo que mientras **más intentos por resolver la evaluación, mayor es el tiempo que**

requiere en algunos casos mientras que para otros tiene menos intentos pero utiliza mayor tiempo para reflexionar sobre las respuestas que se dispone a dar el estudiante. En base a esto se procede a situar al grupo 1 (clúster 0) en un nivel de utilización de la herramienta permanente.

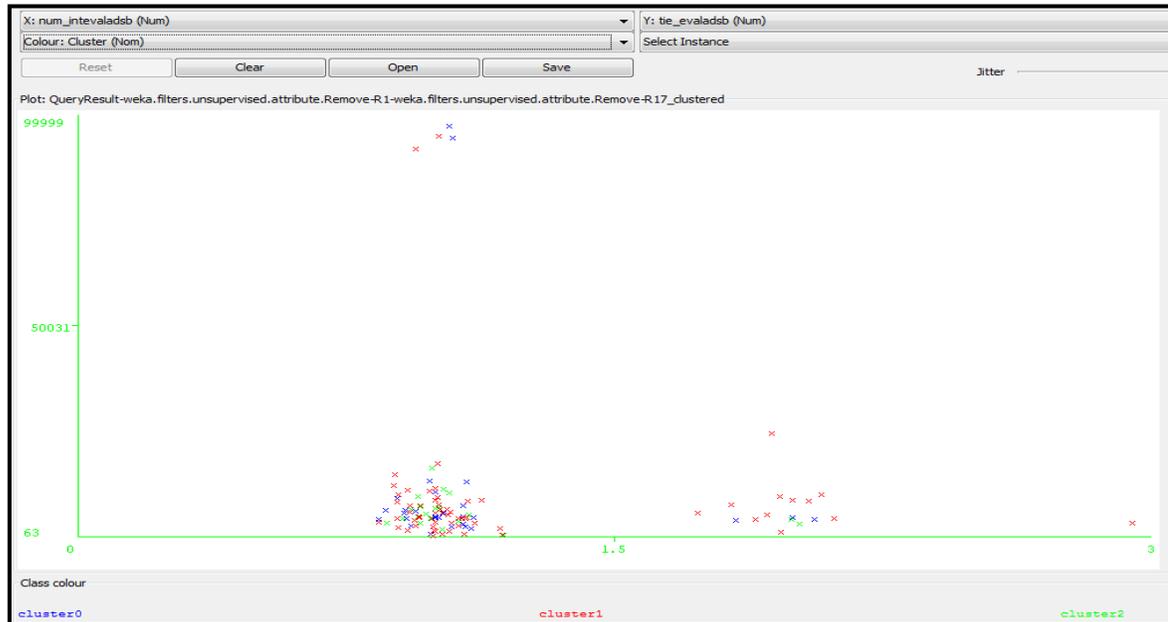


Figura 3.18: Número de intentos por rendir una evaluación según el tiempo utilizado (evaluación a distancia segundo bimestre), (lógica de la programación)

En la figura 3.18 se observa una comparación del tiempo utilizado para rendir la evaluación a distancia del segundo bimestre con el número de intentos por resolver la evaluación encontrando así algunas tendencias en los estudiantes con respecto a la herramienta cuestionario, en la gráfica para el grupo 2 (clúster 1) se nota una instancia que posee **un valor mayor para el tiempo utilizado pero un valor medio para el número de intentos**, pero también se observan varias instancias que pertenecen igual al clúster 1 que tienen los valores medios tanto en el número de intentos por resolver la evaluación como también para los valores del tiempo utilizado.

La herramienta WEKA, tiene la opción de visualizar la asignación de las instancias en cada clúster, aquí ya depende de que se quiera analizar y de esta manera efectuar la combinación de dos atributos que se graficaran y serán sujetos a análisis previo a esto debe existir una correlación entre ambos atributos.

Conclusión de la experimentación con el algoritmo de clustering, los **grupos** que se encontraron al aplicar el algoritmo **SimpleKMeans** son:

- ❖ **Grupo 1** – Clúster 0: Está conformado por los estudiantes que indican mayor cantidad tanto en acceso a los **foros** (42.0833), como en cantidad de contestación a los foros (4.8889) también es mayor que los estudiantes de los demás grupos, lo que significa que conforme tienen *un mayor número de acceso al curso, mayor es la concurrencia para contestar, actualizar, ver o revisar una discusión y agregar un tema de discusión*, en la actualización de la respuesta al foro también es ligeramente mayor en relación al valor del clúster 1, pero con relación al clúster 2 si resulta superior, ahora corresponde mencionar el número de veces que el estudiante agrega un nuevo tema de discusión pues el mayor valor se encuentra en el clúster 0 que es de (1.8611), en cuanto a la descarga de **recursos** conforme el estudiante *accede mayor cantidad de veces al curso* (135.5833), *también presenta valores superiores en la descarga de los mismos* (7.3889) , para el caso de las **tareas** también se encuentra la concentración de los estudiantes que cuando *ingresan o tienen mayor acceso* (135.5833) *al curso, mayor es el número de tareas que envían* (4.5833), en la cuestión de la **mensajería** se encuentran concentrados los estudiantes con el valor más alto tanto para mensajes escritos (3.3494), como para la revisión del historial de mensajes (3.2892), lo que indica que *mientras es mayor el número de mensajes escritos, mayor cantidad de veces realizan la revisión del historial*, en lo que se refiere a **twitter** se muestran los estudiantes que *mientras los valores son altos en el número de accesos al curso* (15.5278), *mayor es el número de mensajes*

*twitter enviados (7.6944), para el cuestionario tiene el mayor número de veces que revisa el cuestionario (17.3333), al igual que el número de intentos por resolverlo (1.1667), ocurriendo esto para la evaluación a distancia del primer bimestre, por otra parte para la evaluación a distancia del segundo bimestre sucede igual, ya que contiene los valores más altos en el número de intentos por resolverlo (0.9167), con un tiempo relativamente mayor (7185.0.938). Por lo tanto se concluye que se trata de un grupo con un nivel de utilización de herramientas (foros, recursos, tareas, mensajería, twitter, cuestionario) ALTO (**permanente**).*

- ❖ **Grupo 2 – Clúster 1:** Formado por los estudiantes que en los **foros** presentan los valores, que mientras *es medio el número de accesos al curso (57.4217), medio es el número de veces que dan contestación (1.7108)*, para la descarga de **recursos** conforme acceden al curso también muestra los valores que se encuentran en un rango medio tanto para *número de descargas (4.4096)* como para *el número de accesos al curso (57.4217)*, en consecuencia denota que, el que un estudiante descargue algún contenido, que habilitó el profesor en cantidad media, depende de los accesos que realice en el curso que refleja igualmente cantidad media, ahora en **tareas** están presentes los estudiantes con un *número de accesos medio al curso (57.4217)* lo que implica que el *número de tareas enviadas sea igual a medio (3.3494)*, para la **mensajería** se identifican los estudiantes que *tienen una cantidad de mensajes escritos medio (3.3494)* implicando que, el *número de veces que revisa el historial de mensajes sea medio también (3.2892)*, para el caso de **twitter** dispone de los estudiantes que han enviado un número de mensajes twitter medio (1.5576) en comparación con el resto de grupos, y en lo que concierne al **cuestionario** este grupo representa una particularidad en relación a los demás grupos y es en el *número de intentos por resolver la evaluación del primer bimestre*, pues presenta el valor medio (1.1446), pero para *el tiempo que el estudiante utiliza para rendir la evaluación* es alto (12767.7756), con un número medio de *revisión del*

*cuestionario o evaluación a distancia del primer bimestre (11.8916) por parte del estudiante, en contraste sucede que para el tiempo que requiere para resolver la evaluación a distancia del segundo bimestre (0.9036) es medio en relación a los demás grupos, aparte de que una vez ya rendida la evaluación los estudiantes de este grupo presentan un valor medio de revisión o visualización del cuestionario (evaluación). Por lo que se puede deducir que el nivel de utilización de herramientas (foros, recursos, tareas, mensajería, twitter, cuestionario) de este grupo es MEDIO (**moderado**).*

- ❖ **Grupo 3 – Clúster 2:** Se encuentran los estudiantes que en los **foros** proceden a *dar contestación con una cantidad baja (0.1512)*, lo que resulta lógico ya que aquí se encuentran los estudiantes que tienen una *cantidad de número de acceso a los foros considerablemente baja (0.1512)*, lo mismo ocurre para el número de veces que agrega un tema de discusión es *baja (0.093)*, encajando con la *cantidad baja (2.4302) del número de veces que revisa una discusión dentro de un foro*, además de que en el *número de veces que actualiza la contestación del foro también es baja (0.1512)*, debiéndose esto a la poca acogida de este grupo de estudiantes en dar contestaciones, para el caso de la descarga de **recursos** tenemos que, por el *número de acceso bajo (16.3605) al curso*, el *número de descargas de los recursos es también bajo (1.6744)*, en las **tareas** se nota que conforme es *bajo (16.3605) el número de accesos al curso de igual manera es bajo (0.3721) el número de veces que suben o envían las tareas*, ahora para la **mensajería** están los estudiantes que *escriben en una cantidad baja (0.9186) los mensajes*, debido a que es *baja (2.407) la cantidad de veces que revisa el historial de mensajes*, pues denota que los estudiantes plantearon pocas interrogantes sobre algún aspecto que necesitarán saber del curso, lo que implica su poca frecuencia para revisar el historial y chequear las respuestas, lo correspondiente a **twitter**, se comprueba que los estudiantes tienen un *número de mensajes generalmente bajo (0.642)*, en discrepancia al resto de grupos, y para el **cuestionario** está conformado por los

estudiantes que tienen un menor acceso para visualizar un cuestionario una vez que ha sido resuelto cuyo valor es (4.1163), pues para algunos estudiantes resulta útil checar en que preguntas fallaron para tomar nota y no volver a cometer los mismos errores, esto tiene relación con el número de accesos al curso cuyo valor es (16.3605), Por lo tanto se concluye que el nivel de utilización de las herramientas (foros, recursos, tareas, mensajería, twitter, cuestionario) de este grupo es BAJO (*escaso*).

### **3.6.2 Experimentación y resultados del curso de ética y derechos humanos**

Para el **primer experimento**:

- En la aplicación WEKA, se escoge el algoritmo clúster SimpleKMeans al momento de realizar la configuración se aplica un valor de 3 clústeres en vista de que se dispone de un número considerable de instancias Ética y derechos humanos (156), para a partir de este buen número de instancias, descubrir grupos definidos de estudiantes, previa una definición de niveles de actividad en la utilización de las herramientas por parte de los estudiantes en la plataforma y estas son Escasa (E) identificado por los valores mínimos, Moderado (M) identificado por los valores medios y Permanente (P) identificado por los valores altos, el resto de valores se deja los que vienen por defecto.

**Tabla 3.22: Primer experimento con el algoritmo SimpleKMeans – semilla 1000 (ética y derechos humanos)**

| Ética y derechos humanos |   |                         |  |  |
|--------------------------|---|-------------------------|--|--|
| Algoritmo                | Clúster   | Instancias clasificadas |  |  |
| SimpleKMeans             | 0   | 40 – 26%                |  |  |
|                          | 1   | 22 – 14%                |  |  |
|                          | 2   | 94 – 60%                |  |  |
|                          | <pre> kMeans *****  Number of iterations: 6 Within cluster sum of squared errors: 33.88927152478818 Missing values globally replaced with mean/mode  Cluster centroids:  Attribute          Full Data      Cluster#                    (156)         0           1           2                    (40)         (22)        (94) ----- view_forum_forum   0.7244         1.525       0.4091      0.4574 add_post_forum     0.0128         0.025       0            0.0106 view_discussion_forum 0.141         0.375       0.0455      0.0638 view_resource      6.0256         6.975       10.1818     4.6489 view_assignment    1.4423         4.65        0.5909      0.2766 upload_upload      0.5962         1.95        0.2727      0.0957 ms_twt             0.2372         0.175       1.3636      0 write_message      2.5513         5.175       3.9545      1.1064 history_message    1.1026         0.85        0.8182      1.2766 view_quiz          1.6987         5.025       2.0909      0.1915 num_intevaladpb   0.3462         1.05        0.2727      0.0638 tie_evaladpb      156.1603       506.775     108.4091    18.1383 num_intevaladsb   0.3077         1.025       0.1818      0.0319 tie_evaladsb      156.3782       531.525     111.5455    7.234 num_accesos_curso 14.4936        23.3        24.4545     8.4149  Clustered Instances  0      40 ( 26%) 1      22 ( 14%) 2      94 ( 60%) </pre> |                         |  |  |

En la tabla 3.22 podemos observar el número de iteraciones que se han efectuado para obtener los clústeres, también tenemos la suma de los errores cuadrados y los porcentajes de instancias clasificadas en cada clúster

Pues en este primer experimento se puede apreciar los resultados obtenidos con el algoritmo SimpleKMeans, teniendo de esta manera en el clúster 1 el 14% de instancias lo

que equivale a 22 y aquí es donde se encuentran establecidas las instancias con los valores medios en relación al clúster 0 y 2, además se puede apreciar la suma de los errores cuadrados de este primer experimento que es de 33.88, pero por tener una mejor variedad de resultados lo que conviene es seguir probando hasta encontrar que las agrupaciones tengan características similares en cuanto a, ***la incidencia de alguna acción que realicen sobre la plataforma.***

Luego de haber realizado algunas experimentaciones se observó que los resultados obtenidos en el primer experimento presentó los mejores resultados, dejando notar que en el clúster 0 se encuentran aquellas instancias con los valores generalmente más altos, tenemos que en el centroide del atributo *view\_resource* (número de veces que descarga un recurso) se encuentra en un valor de 6.975, este valor es mayor al valor que tiene el clúster 1 pero menor que el del clúster 1, encontrando esta excepción para el atributo *view\_resource* que presenta el valor medio entre el clúster 1 y 2, además que en el atributo *num\_accesos\_curso* (número de accesos al curso) también se presenta en las mismas condiciones que el atributo *view\_resource*, es decir ubicado en un nivel medio en relación al clúster 1 y 2 teniendo el clúster 0 el valor de 23.3 en *num\_accesos\_curso*, con esto se puede concluir que por lo general los alumnos de la asignatura de *ética y derechos humanos* de la carrera de *abogacía* con un número medio de accesos al curso, descargan los recursos de la plataforma en una incidencia media, es decir es relativo el número de descargas con el número de accesos al curso y se presenta en iguales condiciones de uso ***moderado*** en cuanto al nivel de uso de la herramienta ***recursos***.

Otro detalle que cabe indicar es el valor del atributo *ms\_twt* (número de mensajes twitter) en el clúster 0 pues es menor que el clúster 1, pero mayor que el clúster 2, notándose que está en un valor medio, este atributo también es relativo al atributo *num\_accesos\_curso* que se encuentra también en un valor medio en comparación al clúster 1 y 2, lo que indica que los estudiantes con un número medio de accesos al curso

envían un número medio de mensajes twitter, con esto podemos derivar que los estudiantes presentan un uso **moderado** que corresponde al nivel de uso de la herramienta **twitter**. El resto de atributos clasificados en el clúster 0 poseen los valores más altos nivel de utilización de la herramienta **permanente**, en relación a los demás clúster ya que en el clúster 1 se encuentran clasificados la mayoría de valores medios nivel de utilización de la herramienta **moderado**, y en el clúster 2 los valores mínimos nivel de utilización de la herramienta **escaso**.

Para el **segundo experimento**:

- Aquí se mantiene el mismo número de clúster y los valores restantes que están por defecto, solo que esta vez se aumentará el valor de seed (semilla) pasaremos de 1000 con el que se probó en un inicio a 3000, esperando se minimice la suma del error cuadrático y exista homogeneidad entre los grupos. En la tabla 3.23 se puede apreciar los resultados de un segundo experimento.

*Tabla 3.23: Segundo experimento con el algoritmo SimpleKMeans – semilla 3000 (ética y derechos humanos)*

| Ética y derechos humanos |   |                         |  |
|--------------------------|---|-------------------------|--|
| Algoritmo                | Clúster   | Instancias clasificadas |  |
| SimpleKMeans             | 0   | 38 – 24%                |  |
|                          | 1   | 99 – 63%                |  |
|                          | 2   | 19 – 12%                |  |
|                          | <pre> kMeans *****  Number of iterations: 8 Within cluster sum of squared errors: 38.01866954653754 Missing values globally replaced with mean/mode  Cluster centroids:  Attribute          Full Data      Cluster#                    (156)         0           1           2                    (38)         (99)        (19) ----- view_forum_forum   0.7244         1.4474      0.4949      0.4737 add_post_forum     0.0128         0.0263      0.0101      0 view_discussion_forum 0.141         0.3947      0.0707      0 view_resource      6.0256         7.0526      5.5859      6.2632 view_assignment    1.4423         4.7895      0.0808      1.8421 upload_upload      0.5962         2.0263      0            0.8421 ms_twt             0.2372         0.2895      0.202       0.3158 write_message      2.5513         5.2895      1.1919      4.1579 history_message    1.1026         0.8158      0.303       5.8421 view_quiz          1.6987         5.3158      0.0303      3.1579 num_intevaladpb    0.3462         1.2105      0            0.4211 tie_evaladpb       156.1603      547.6316    0            186.8947 num_intevaladsb    0.3077         0.9737      0            0.5789 tie_evaladsb       156.3782      534.8421    0            214.2632 num_accesos_curso  14.4936        24.1316    10.3232     16.9474  Clustered Instances  0      38 ( 24%) 1      99 ( 63%) 2      19 ( 12%) </pre> |                         |  |

Esta vez se presenta una agrupación en la que el clúster 1 posee el valor de 63% pues es notable que tiene el mayor valor respecto al clúster 0 y 2, en el caso del clúster 2 este tiene el menor valor de instancias clasificadas que es del 12% en relación al clúster 0 y 1, además de que en el clúster 0 se agrupan las instancias con valores alto, pero si nos fijamos en el clúster 2 podemos apreciar que en algunos atributos tiene valores más altos que el clúster 1, lo que da como resultado una agrupación ambigua ya que también

en este experimento se nota un aumento en la suma del error cuadrático que es de 38.01 es claro que no es confiable ya que implica que no se realizó una óptima definición de los grupos lo que conlleva a realizar otro experimento.

Para el **tercer experimento**:

- Aquí se realizó algunos experimentos en cuanto a la configuración del valor de la semilla, existiendo un caso muy particular en la suma del error cuadrático pues es prácticamente el mismo que el caso del primer experimento el cual fue de 33.88 difiriendo con dos decimas al tercer experimento que es de 33.90, esto originó que se hicieran más pruebas pero esta vez con un valor de semilla más alto pues a partir del rango 8000 de semilla la suma del error cuadrático aumentaba considerablemente por estas razones se estima conveniente seleccionar los resultados del **primer experimento** de la tabla 3.22 para proceder a la interpretación y análisis de los resultados.

*Tabla 3.24: Tercer experimento con el algoritmo SimpleKMeans – semilla 7000 (ética y derechos humanos)*

| Ética y derechos humanos |  |                         |  |  |
|--------------------------|--|-------------------------|--|--|
| Algoritmo                | Clúster  | Instancias clasificadas |  |  |
| SimpleKMeans             | 0  | 22 – 14%                |  |  |
|                          | 1  | 91 – 58%                |  |  |
|                          | 2  | 43 – 28%                |  |  |
|                          | <pre> kMeans *****  Number of iterations: 5 Within cluster sum of squared errors: 33.90001111661718 Missing values globally replaced with mean/mode  Cluster centroids:  Attribute          Full Data      Cluster#                    (156)         0           1           2                    (22)         (91)        (43) ----- view_forum_forum   0.7244         0.4091      0.4725      1.4186 add_post_forum     0.0128         0           0.011       0.0233 view_discussion_forum 0.141         0.0455      0.0659      0.3488 view_resource      6.0256        10.1818     4.6813      6.7442 view_assignment    1.4423         0.5909      0.2198      4.4651 upload_upload      0.5962         0.2727      0.0659      1.8837 ms_twt             0.2372         1.3636      0           0.1628 write_message      2.5513         3.9545      1.1099      4.8837 history_message    1.1026         0.8182      1.3187      0.7907 view_quiz          1.6987         2.0909      0.1319      4.814 num_intevaladpb   0.3462         0.2727      0.0659      0.9767 tie_evaladpb      156.1603      108.4091    18.7363     471.4186 num_intevaladsb   0.3077         0.1818      0           1.0233 tie_evaladsb      156.3782      111.5455    0           510.2558 num_accesos_curso 14.4936        24.4545     8.4176      22.2558  Clustered Instances  0      22 ( 14%) 1      91 ( 58%) 2      43 ( 28%) </pre> |                         |  |  |

A continuación se evalúan los resultados de los tres experimentos realizados para de esta manera establecer cuál de éstos es el que da mejores resultados.

*Tabla 3.25: Comparativa de los experimentos realizados utilizando SimpleKMeans (ética y derechos humanos)*

| SimpleKMeans               | Clúster | Instancias Clasificadas | Suma de errores cuadrados |
|----------------------------|---------|-------------------------|---------------------------|
| <b>Primer Experimento</b>  | 0       | 40 – 26%                | 33.88                     |
|                            | 1       | 22 – 14%                |                           |
|                            | 2       | 94 – 60%                |                           |
| <b>Segundo Experimento</b> | 0       | 38 – 24%                | 38.01                     |
|                            | 1       | 99 – 63%                |                           |
|                            | 2       | 19 – 12%                |                           |
| <b>Tercer Experimento</b>  | 0       | 22 – 14%                | 33.90                     |
|                            | 1       | 91 – 58%                |                           |
|                            | 2       | 43 – 28%                |                           |

En la tabla 3.25 observamos los resultados de cada experimento, encontrando que los valores obtenidos del primer experimento presenta resultados aceptables entre el clúster 0, 1 y 2 pero en la suma de errores cuadrados es el que tiene el valor menor de todos los experimentos pero pese a eso se siguió realizando pruebas para indagar que pudiesen haber más alternativas que superen en optimización sus resultados, en el segundo experimento no se encuentra una agrupación semejante entre los clúster por esto fue desechado además de ser el que posee un mayor valor en la suma de errores cuadrados y se optó por seguir probando llegando así a un tercer experimento aquí se contó con una distribución semejante al del primer experimento de las instancias en la agrupación pero con una diferencia no tan marcada en cuanto a la suma de los errores cuadrados, respecto al experimento primero reflejando este ser el más adecuado para hacer la interpretación acorde con los resultados basados en los grupos.

- ❖ **Utilización de herramientas**, para este indicador se considera los siguientes atributos en cada clúster, aquí se muestra el análisis para obtener el nivel de utilización de las herramientas por parte de los estudiantes:

**Tabla 3.26: Nivel de utilización de las herramientas (ética y derechos humanos)**

| Herramienta  | Clúster | Atributo                     | Centroide | Nivel utilización herramienta |
|--------------|---------|------------------------------|-----------|-------------------------------|
| Foros        | 0       | <i>view_forum_forum</i>      | 1.525     | Permanente                    |
|              |         | <i>add_post_forum</i>        | 0.025     |                               |
|              |         | <i>view_discussion_forum</i> | 0.375     |                               |
|              | 1       | <i>view_forum_forum</i>      | 0.4091    | Escaso                        |
|              |         | <i>add_post_forum</i>        | 0         |                               |
|              |         | <i>view_discussion_forum</i> | 0.0455    |                               |
|              | 2       | <i>view_forum_forum</i>      | 0.4574    | Moderado                      |
|              |         | <i>add_post_forum</i>        | 0.0106    |                               |
|              |         | <i>view_discussion_forum</i> | 0.0638    |                               |
| Recursos     | 0       | <i>view_resource</i>         | 6.975     | Moderado                      |
|              | 1       | <i>view_resource</i>         | 10.1818   | Permanente                    |
|              | 2       | <i>view_resource</i>         | 4.6489    | Escaso                        |
| Tareas       | 0       | <i>view_assignment</i>       | 4.65      | Permanente                    |
|              |         | <i>upload_upload</i>         | 1.95      |                               |
|              | 1       | <i>view_assignment</i>       | 0.5909    | Moderado                      |
|              |         | <i>upload_upload</i>         | 0.2727    |                               |
|              | 2       | <i>view_assignment</i>       | 0.2766    | Escaso                        |
|              |         | <i>upload_upload</i>         | 0.0957    |                               |
| Mensajería   | 0       | <i>write_message</i>         | 5.175     | Permanente                    |
|              |         | <i>history_message</i>       | 0.85      |                               |
|              | 1       | <i>write_message</i>         | 3.9545    | Moderado                      |
|              |         | <i>history_message</i>       | 0.8182    |                               |
|              | 2       | <i>write_message</i>         | 1.1064    | Escaso                        |
|              |         | <i>history_message</i>       | 1.2766    |                               |
| Twitter      | 0       | <i>ms_twitter</i>            | 0.175     | Moderado                      |
|              | 1       | <i>ms_twitter</i>            | 1.3636    | Permanente                    |
|              | 2       | <i>ms_twitter</i>            | 0         | Escaso                        |
| Cuestionario | 0       | <i>num_intevaldpb</i>        | 1.05      | Permanente                    |
|              |         | <i>tie_evaladpb</i>          | 506.775   |                               |
|              |         | <i>num_intevaladsb</i>       | 1.025     |                               |
|              |         | <i>tie_evaladsb</i>          | 531.525   |                               |
|              |         | <i>view_quiz</i>             | 5.025     |                               |
|              |         | <i>num_intevaldpb</i>        | 0.2727    |                               |
|              |         | <i>tie_evaladpb</i>          | 108.4091  |                               |
|              |         | <i>num_intevaladsb</i>       | 0.1818    |                               |

|  |   |                        |          |          |
|--|---|------------------------|----------|----------|
|  | 1 | <i>tie_evaladsb</i>    | 111.2588 | Moderado |
|  |   | <i>view_quiz</i>       | 2.0909   |          |
|  | 2 | <i>num_intevaldpb</i>  | 0.0638   | Escaso   |
|  |   | <i>tie_evaladpb</i>    | 18.1383  |          |
|  |   | <i>num_intevaladsb</i> | 0.0319   |          |
|  |   | <i>tie_evaladsb</i>    | 7.234    |          |
|  |   | <i>view_quiz</i>       | 0.1915   |          |

### FOROS

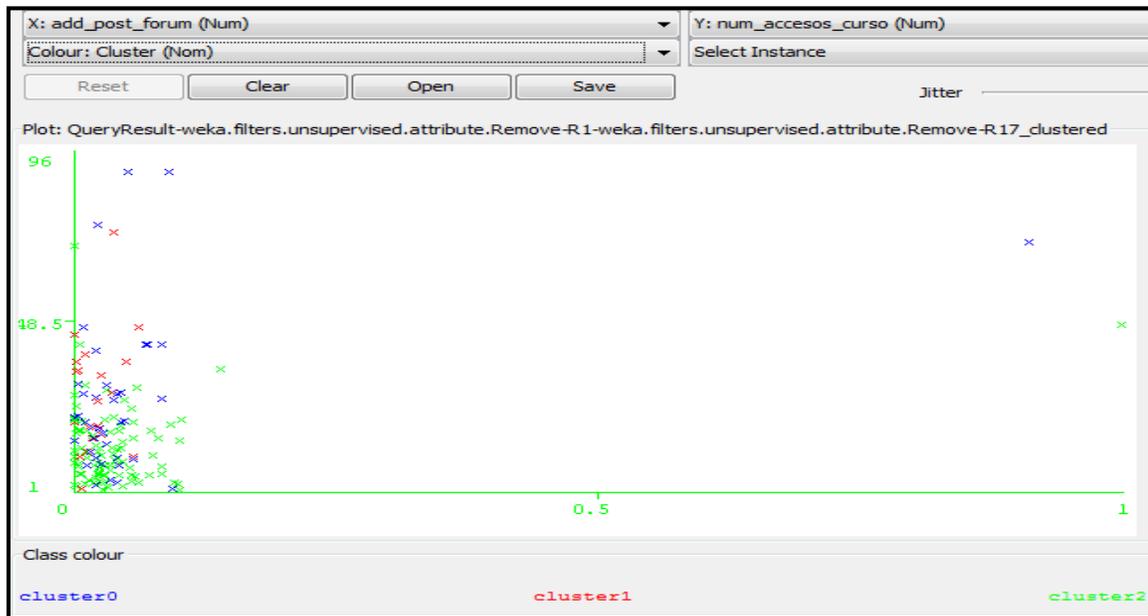


Figura 3.19: Número de veces que contesta un foro según el número de accesos al curso (ética y derechos humanos)

En la figura 3.19 se observa que el número de veces que contestan un foro los estudiantes está ubicado en el eje x y el número de accesos al curso en el eje de la y, notándose cada clúster con su respectivo color para diferenciar las instancias que le corresponden a cada uno.

En esta gráfica se ve la distribución de los tres grupos en cada atributo, se puede notar que en el grupo 1 (clúster 0) en el atributo *add\_post\_forum* (número de veces que contesta un foro) presenta un patrón, que mientras **más alto es el número de accesos al curso, mayor**

**(permanente)** es el número de veces que contestan los estudiantes, el grupo 2 (clúster 1) se aprecia que se encuentran concentrados los estudiantes que mientras han tenido acceso a foros bajo (*escaso*), lo mismo ocurre con la contestación al foro es bajo, y en el grupo 3 (clúster 2) se nota que se encuentran ubicados los estudiantes que tienen un número de contestación medio (*moderado*), lo que denota y se aprecia en la grafica que en número de veces que contesta es medio.

### RECURSOS

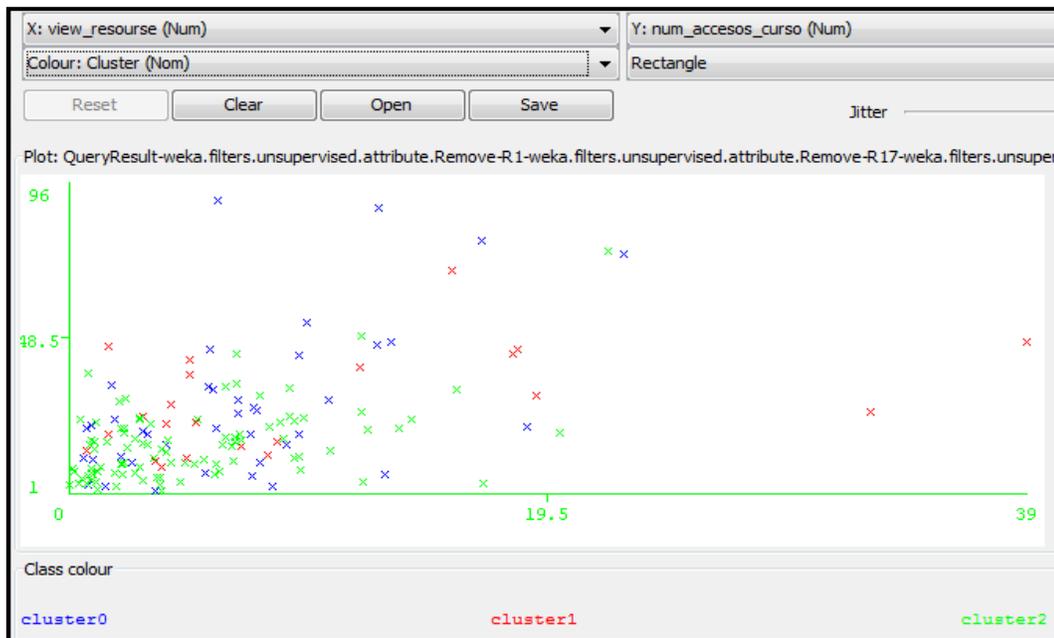


Figura 3.20: Número de veces que descarga un recurso según el número de accesos al curso (ética y derechos humanos)

Como se puede ver en la figura 3.20 en el eje de la x se encuentra representado el atributo *view\_resource* (número de veces que descarga un recurso), mientras que en el eje de y el número de accesos al curso, y cada una de las instancias de los atributos coloreadas con el clúster que les compete.

En la gráfica se indica que en el grupo 2 (clúster 1) están situados los estudiantes que conforme el número de accesos al curso ha sido alto (*permanente*), mayor es la cantidad de veces que descarga un recurso, en el grupo 1 se encuentran concentrados los estudiantes que presentan un *nivel de utilización de la herramienta* medio (*moderado*), y en el caso del grupo 3 los estudiantes con un *nivel de utilización de la herramienta* bajo (*escaso*).

### TAREAS

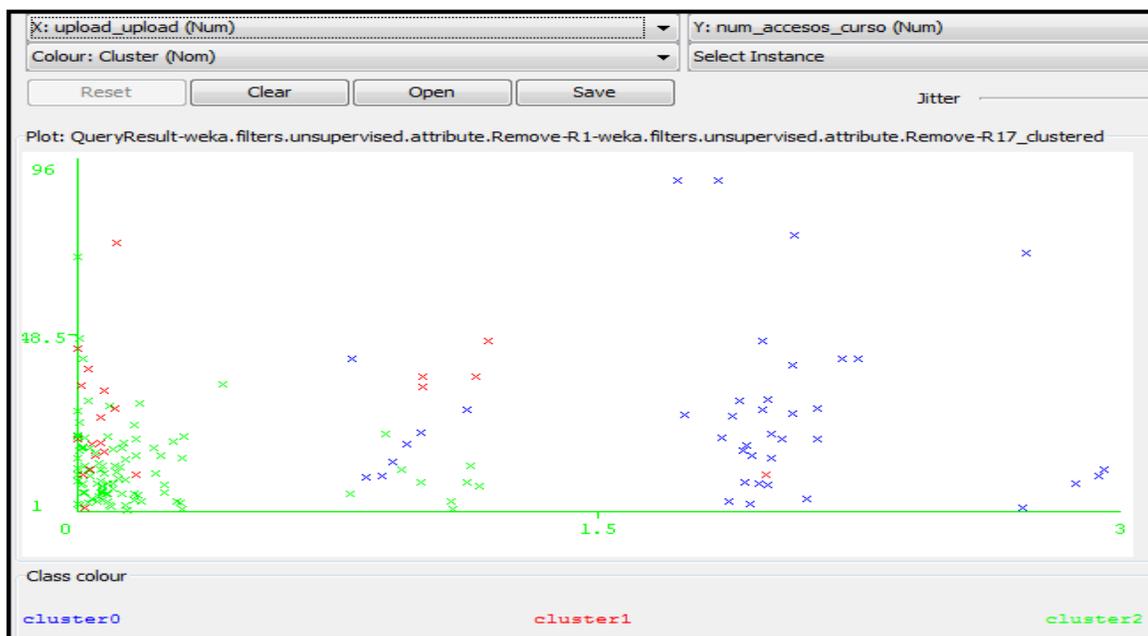


Figura 3.21: Número de veces que envía o sube una tarea según número de accesos al curso (ética y derechos humanos)

En la figura 3.21 se puede observar el atributo *upload\_upload* (número de veces que sube una tarea) está ubicado en el eje de la *x*, y el atributo número de accesos curso en el eje de la *y*.

En la gráfica se tiene que en el grupo 1 (clúster 0), se manifiesta una mayor concentración de estudiantes con los valores más altos, que acceden a las tareas para verificar la existencia de estas y la fecha límite para ser enviadas, lo que demuestra que el número de tareas enviadas también es alto mostrando la tendencia de este grupo 1 que mientras **mayor es el acceso al**

curso para revisar las tareas enviadas por el profesor, mayor es el número de tareas enviadas lo que implica que tienen un nivel de utilización de la herramienta alto (*permanente*), en el grupo 2 (clúster 1) se ubican los que poseen un nivel medio (*moderado*) y en el grupo 3 (clúster 2) están presentes los estudiantes con un nivel bajo (*escaso*).

### MENSAJERÍA

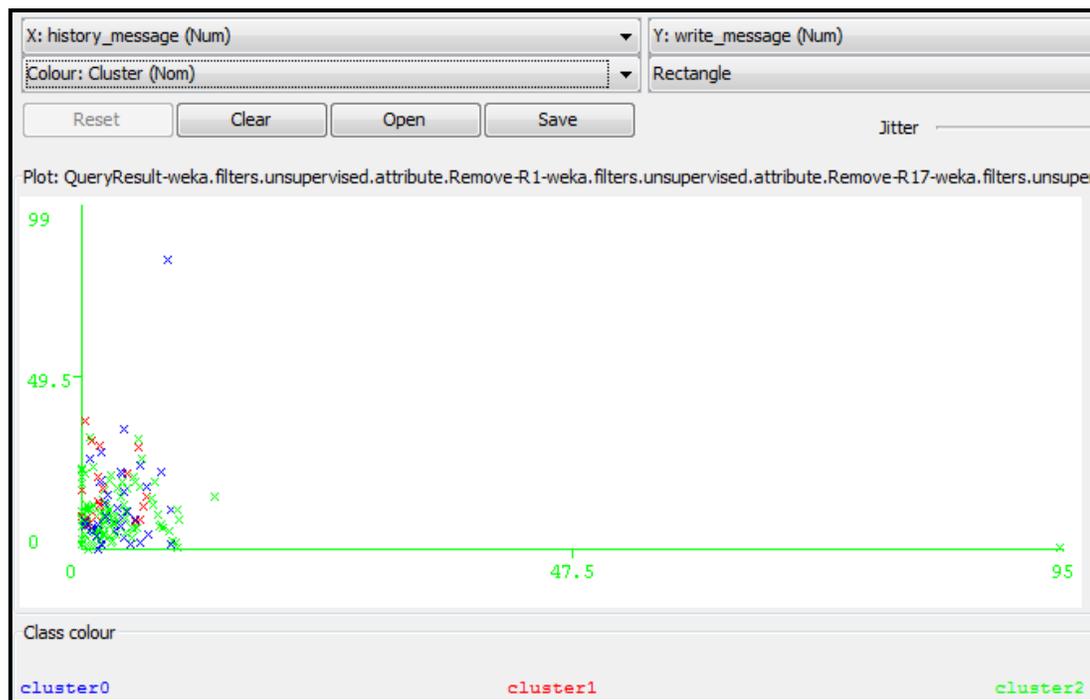


Figura 3.22: Número de veces que revisa el historial de mensajes según número de mensajes que ha escrito (ética y derechos humanos)

En la figura 3.22 observamos en el eje de la x el atributo *history\_message* (número de veces que revisa el historial de mensajes), y en el eje de la y *write\_message* (número de mensajes que ha escrito).

En la gráfica se aprecia tres grupos partiremos con el grupo 1 (clúster 0) aquí están presentes los estudiantes que poseen los valores más altos en cuanto al número de mensajes escritos pero, no siendo así en el caso del número de veces que revisa el historial de mensajes aquí

denota los valores en un nivel medio, conviene aclarar que éstos atributos están interrelacionados y por el hecho de que el número de mensajes escritos presenta los valores más altos en este grupo 1 se presume que éste grupo tiene un nivel de utilización de la herramienta (*permanente*), acentuando que en los estudiantes necesitan en una escala media revisar el historial de mensajes y leer las posibles contestaciones a alguna inquietud a su profesor o compañeros, el grupo 3 (clúster 2) en cambio presenta los valores más altos en el caso de la revisión del historial de mensajes pero con un número de mensajes escritos bajo por lo que se concluye que los estudiantes de este grupo presentan un nivel de utilización de la herramienta *escaso*, lo que demuestra que aquí los estudiantes tienden a estar constantemente revisando su historial para observar algún comunicado por parte del profesor pero sus inquietudes son realmente inferiores en relación a los demás grupos, y en el grupo 1 se tiene a los estudiantes con un nivel de utilización de la herramienta alto (*permanente*).

### TWITTER

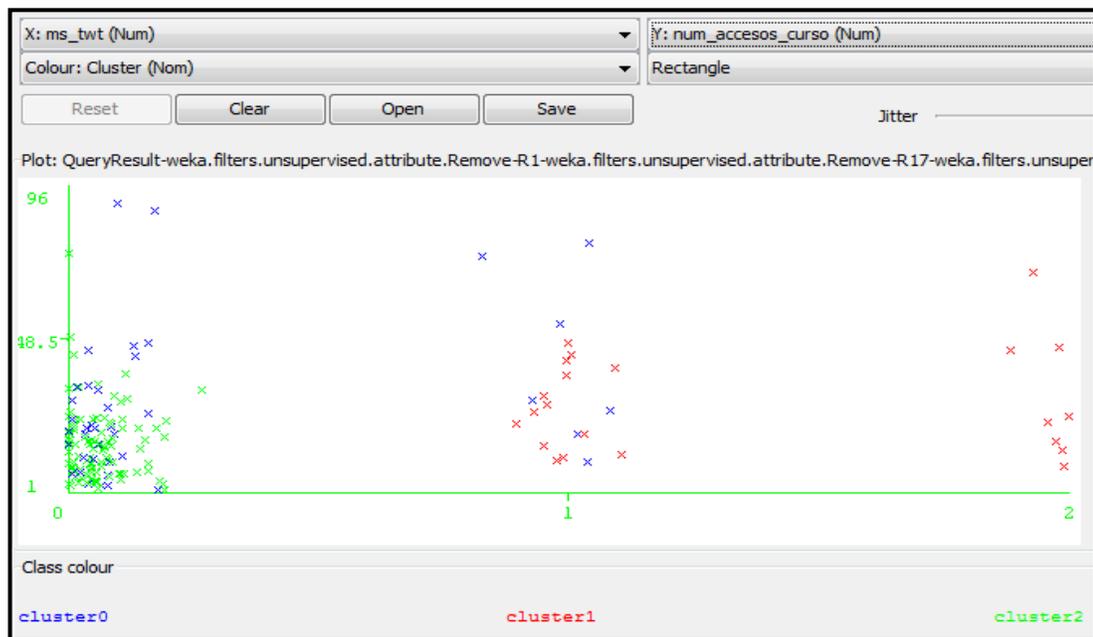


Figura 3.23: Número de mensajes twitter según el número de accesos al curso (ética y derechos humanos)

En la figura 3.23 se aprecia para el atributo *ms\_twt* (número de mensajes twitter) en el eje de la x, mientras que el atributo del número de accesos al curso en el eje de la y.

En la gráfica denota tres grupos definidos para el caso del grupo 2 (clúster 1) se encuentran los estudiantes **que mientras mayor es el número de accesos al curso, mayor es el número de mensajes twitter** lo que denota que en este grupo incide el nivel de utilización de la herramienta (*Permanente*) lo que resulta lógico ya que conforme los estudiantes acceden al curso manifiestan alguna interrogante sobre alguna tarea, recurso, contenido o evaluaciones de la asignatura por medio de un mensaje twitter de manera pública, es decir a vista de todos los participantes del curso para que el primero que lo observe o le preste atención proceda a contestar, ahora corresponde analizar los estudiantes que intervienen en el grupo 1 (clúster 0) pues son los que denotan los valores medios tanto en el número de mensajes twitter como en el número de accesos al curso, lo que conlleva a concluir que tienen un nivel de utilización de la herramienta *moderado*, el grupo 3 presenta los estudiantes que es *escasa* la utilización de la herramienta.

### CUESTIONARIO

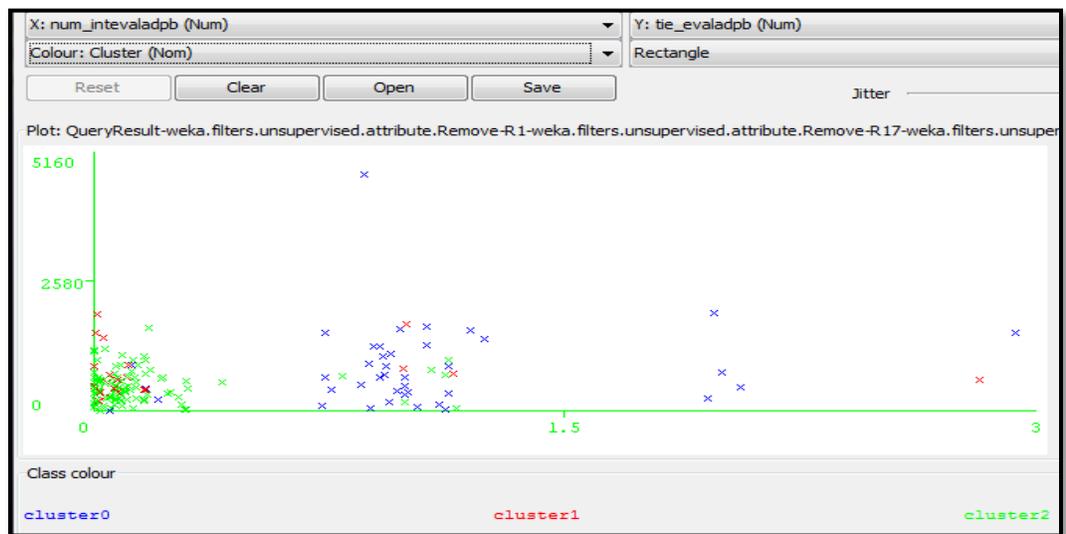


Figura 3.24: Número de intentos por rendir una evaluación según el tiempo utilizado (evaluación a distancia primer bimestre), (ética y derechos humanos)

En la figura 3.24 se observa para el atributo *num\_intevaladpb* (número de intentos en la evaluación a distancia del primer bimestre) en el eje de la *x*, en lo que corresponde al atributo *tie\_evaladpb* (tiempo que tarda en la evaluación a distancia del primer bimestre) en el eje de la *y*.

En la gráfica se observa una asimilación del tiempo utilizado para rendir la evaluación a distancia del primer bimestre con el número de intentos por resolver la evaluación, encontrando que en el grupo 1 (clúster 0) se encuentran los estudiantes que mientras **es alto el tiempo que necesitan para rendir la evaluación a distancia del primer bimestre, mayor el número de intentos** lo que indica que se trata de los estudiantes que leen a la ligera las preguntas o las dejan en blanco y como tienen oportunidad de un nuevo intento para mejorar o completar las preguntas que no fueron contestadas con anterioridad realizan un mayor número de intentos lo que implica que el tiempo se vaya acumulando y sea alto, en base a esto se deriva que éste grupo presenta un nivel de utilización de la herramienta *permanente*, el grupo 2 (clúster 1) tiene la concentración de los estudiantes que poseen valores medios en ambos atributos por lo que expresa que utilizan la herramienta en un nivel *moderado* y en el grupo 3 (clúster 2) los estudiantes que presentan los valores bajos indicando que utilizan la herramienta en un nivel *escaso*.

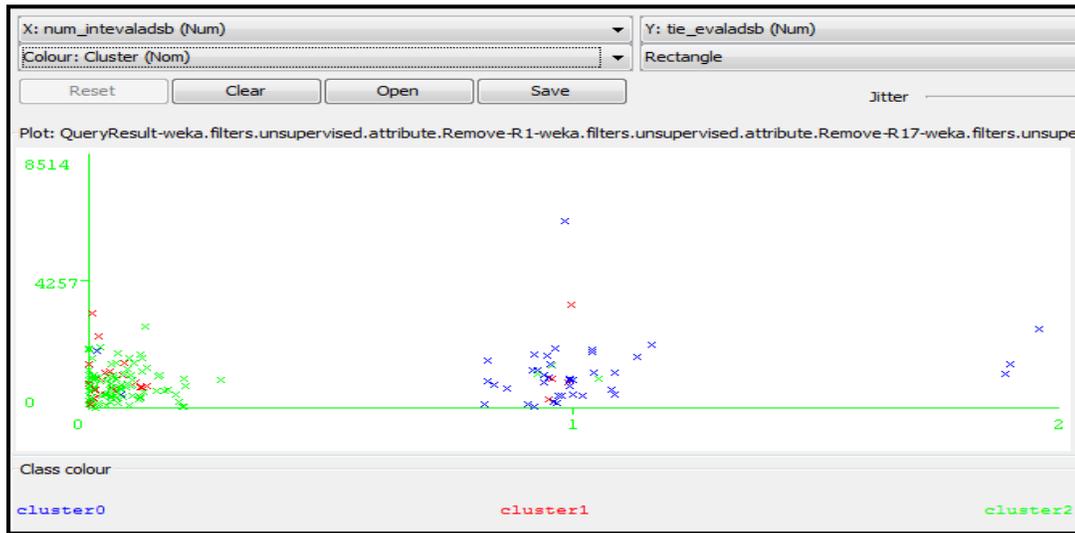


Figura 3.25: Número de intentos por rendir una evaluación según el tiempo utilizado (evaluación a distancia segundo bimestre), (ética y derechos humanos)

Como se puede apreciar en la figura 3.25 en el eje de la x se tiene el atributo *num\_intevaladsb* (número de intentos en la evaluación a distancia del segundo bimestre), y para el atributo *tie\_evaladsb* (tiempo que tarda en la evaluación a distancia del segundo bimestre) en el eje de la y.

En la gráfica se aprecia tres grupos cada uno con su respectiva tendencia de los estudiantes en cuanto al número de intentos según el tiempo que utilizan por rendir la evaluación a distancia del segundo bimestre, se nota en el grupo 1 (clúster 0) el acopio de los estudiantes que **realizan el mayor número de intentos por resolver la evaluación, tienden a requerir de más tiempo** lo que demuestra que son los estudiantes que tienen mayores equivocaciones en las respuestas de la evaluación, lo que los obliga a realizar más intentos involucrando un aumento en el tiempo que utilizan por lo que se concluye que en este grupo se presenta un nivel de utilización de la herramienta *permanente*, en tanto el nivel de utilización de la herramienta *moderado* está representado por el grupo 2 (clúster 1) ya que la concentración de los estudiantes son los que corresponden a los valores medios, y el nivel de utilización de la herramienta *escaso* el grupo 3 (clúster 2) los estudiantes que corresponden a los valores bajo.

Conclusión de la experimentación con el algoritmo de clustering, los **grupos** que se encontraron después de aplicar el algoritmo **SimpleKMeans** son:

- ❖ **Grupo 1** – Clúster 0: Está representado por los estudiantes que tienen la mayor cantidad de acceso a los **foros** (1.525), así como en la cantidad de contestación (0.025) y el número de veces que revisa la discusión dentro del foro (0.375), también es mayor que los estudiantes de los demás grupos, lo que significa que conforme tienen *un mayor número de acceso al curso, mayor es la concurrencia para contestar*, en cuanto a la descarga de **recursos** conforme el estudiante *accede en cantidad de veces media al curso (23.3), presenta valores medios en la descarga de los mismos (6.975)*, para el caso de las **tareas** se encuentra la concentración de los estudiantes que cuando *ingresan o tienen mayor acceso al curso, mayor es el número de tareas que envían (1.95)*, en la cuestión de la **mensajería** se encuentran concentrados los estudiantes con el valor más alto tanto para mensajes escritos (5.175), con una revisión media del historial de mensajes (0.85), lo que indica que *mientras es mayor el número de mensajes escritos, media es cantidad de veces para la revisión del historial*, en lo que se refiere a **twitter** se muestran los estudiantes que mientras *los valores son medios en el número de accesos al curso, medio es el número de mensajes twitter enviados (0.175)*, para el **cuestionario** tiene el *mayor número de veces que revisa el cuestionario (5.025), al igual que el número de intentos por resolverlo (1.05)*, ocurriendo esto para la evaluación a distancia del primer bimestre, por otra parte para la evaluación a distancia del segundo bimestre sucede igual, ya que contiene los valores más altos en el *número de intentos por resolverlo (1.025)*, con un *tiempo relativamente mayor (531.525)*. Por lo tanto se concluye que se trata de un grupo con un nivel de utilización de herramientas (foros, recursos, tareas, mensajería, y cuestionario) ALTO (**permanente**). existiendo una excepción para la utilización de la herramienta (recursos y twitter) ya que éste grupo presenta un nivel MEDIO (**moderado**).

- ❖ **Grupo 2** – Clúster 1: Formado por los estudiantes que en los **foros** presentan los valores, que mientras *es bajo el número de acceso a éstos (0.4091)*, *bajo es el número de veces que dan contestación*, para la descarga de **recursos** conforme *acceden al curso* en una cantidad alta (24.45), muestra los valores que se encuentran en un rango alto tanto para *número de descargas (10.1818)*, en consecuencia denota que, las descargas en cantidad alta que hace el estudiante de algún contenido, habilitado por el profesor, depende de los accesos que realice en el curso que refleja igualmente cantidad alta, ahora en **tareas** están presentes los estudiantes con un *número de tareas enviadas igual a medio (0.2727)*, para la **mensajería** se identifican los estudiantes que *tienen una cantidad de mensajes escritos medio (3.9545)* implicando que, el *número de veces que revisa el historial de mensajes sea medio también (0.8182)*, para el caso de **twitter** dispone de los estudiantes que han enviado un número de mensajes twitter alto (1.3636) en comparación con el resto de grupos , y en lo que concierne al **cuestionario** este grupo representa en el *número de intentos por resolver la evaluación del primer bimestre*, el valor medio (0.2727), por consiguiente para *el tiempo que el estudiante utiliza para rendir la evaluación es medio (108.4091)*, con un número medio de *revisión media del cuestionario o evaluación a distancia del primer bimestre (2.0909)* por parte del estudiante, y para *la evaluación a distancia del segundo bimestre el tiempo que requiere para resolver es medio (111.2588)*, sucediendo lo mismo para el número de intentos en la evaluación a distancia del segundo bimestre es medio (0.1818). En base a todas estas observaciones, se puede deducir que el nivel de utilización de herramientas (tareas, mensajería, y cuestionario) de este grupo es MEDIO (**moderado**), existiendo algunas excepciones para la herramienta (recursos y twitter) que presentan un nivel de incidencia ALTO (**permanente**), y para la herramienta (foros) muestra un nivel de utilización BAJO (**escaso**).

- ❖ **Grupo 3 – Clúster 2:** Se encuentran los estudiantes que en los **foros** proceden a *dar contestación con una cantidad media* (0.0106), lo que resulta lógico ya que aquí se encuentran los estudiantes que tienen una *cantidad de número de acceso a los foros* considerablemente media (0.4574), para el caso de la descarga de **recursos** tenemos que, por el *número de acceso* bajo (8.4149) al curso, el *número de descargas* de los recursos es también bajo (4.6489), en las **tareas** se nota que conforme es bajo el *número de accesos* al curso de igual manera es bajo (0.0957) el *número de veces que suben o envían las tareas*, ahora para la **mensajería** están los estudiantes que *escriben en una cantidad baja* (1.1064) los mensajes, debido a que es baja (1.2766) la *cantidad de veces que revisa el historial de mensajes*, pues denota que los estudiantes plantearon pocas interrogantes sobre algún aspecto que necesitaran saber del curso, lo que implica su poca frecuencia para revisar el historial y chequear las respuestas, lo correspondiente a **twitter**, se comprueba que los estudiantes tienen un *número de mensajes* generalmente bajo, igual a (0), en discrepancia al resto de grupos, y para el **cuestionario** está conformado por los estudiantes que tienen un bajo acceso para visualizar un cuestionario (0.1915), lo que resulta comprensible ya que esta acción la realizan una vez que ha sido resuelto, con el objetivo de visualizar sus puntuaciones en las respuestas, esto tiene relación con el número relativamente bajo, de intentos por resolver las evaluaciones a distancia tanto del primer bimestre como del segundo. Por lo tanto a partir de estas particularidades se concluye que el nivel de utilización de las herramientas (recursos, tareas, mensajería, twitter, cuestionario) de este grupo es **BAJO (escaso)**, a excepción de la herramienta (foros) ya que indica un nivel de utilización **MEDIO (moderado)**.

### 3.6.3 Comparación de los grupos entre los cursos: lógica de la programación y, ética y derechos humanos obtenidos con el algoritmo de clustering *SimpleKMeans*

*Tabla 3.27: Comparación de grupos*

| Nivel de utilización de las herramientas |                |                           |            |                          |                   |
|--|----------------|---------------------------|------------|--------------------------|-------------------|
| Herramienta                              | Grupos Clúster | Lógica de la programación |            | Ética y derechos humanos |                   |
|  |                | Instancias Clasificadas   | Nivel      | Instancias Clasificadas  | Nivel             |
| Foros                                    | 0              | 18%                       | Permanente | 26%                      | Permanente        |
|  | 1              | 40%                       | Moderado   | 14%                      | <b>Escaso</b>     |
|  | 2              | 42%                       | Escaso     | 60%                      | <b>Moderado</b>   |
| Recursos                                 | 0              | 18%                       | Permanente | 26%                      | <b>Moderado</b>   |
|  | 1              | 40%                       | Moderado   | 14%                      | <b>Permanente</b> |
|  | 2              | 42%                       | Escaso     | 60%                      | Escaso            |
| Tareas                                   | 0              | 18%                       | Permanente | 26%                      | Permanente        |
|  | 1              | 40%                       | Moderado   | 14%                      | Moderado          |
|  | 2              | 42%                       | Escaso     | 60%                      | Escaso            |
| Mensajería                               | 0              | 18%                       | Permanente | 26%                      | Permanente        |
|  | 1              | 40%                       | Moderado   | 14%                      | Moderado          |
|  | 2              | 42%                       | Escaso     | 60%                      | Escaso            |
| Twitter                                  | 0              | 18%                       | Permanente | 26%                      | <b>Moderado</b>   |
|  | 1              | 40%                       | Moderado   | 14%                      | <b>Permanente</b> |
|  | 2              | 42%                       | Escaso     | 60%                      | Escaso            |
| Cuestionario                             | 0              | 18%                       | Permanente | 26%                      | Permanente        |
|  | 1              | 40%                       | Moderado   | 14%                      | Moderado          |
|  | 2              | 42%                       | Escaso     | 60%                      | Escaso            |

En la tabla 3.27 se muestra la comparación de los resultados entre un curso de la carrera de informática *lógica de la programación* y un curso de la carrera de abogacía *ética y derechos humanos*, acerca del nivel de utilización de las herramientas con el objetivo de descubrir cuál de las herramientas tiene mayor acogida tanto en los estudiantes de una carrera técnica, como socio humanística, y se han encontrado las particularidades que se describen a continuación:

- ❖ **Foros:** En el curso de *lógica de la programación*, se observa el grupo 1 (clúster 0) está conformado por el 18% de los estudiantes que tienen mayor tendencia de la utilización de la herramienta en un nivel *permanente*, en el grupo 2 (clúster 1) con un 40% *moderado*, mientras que para el curso de *ética y derechos humanos*, el grupo 1 (clúster 0) el 26% de los estudiantes presenta nivel *permanente* y el grupo 3 (clúster 2) un 60% nivel *moderado*.
- ❖ **Recursos:** Esta herramienta en el grupo 1 (clúster 0), del curso de *lógica de la programación* con un 18% de los estudiantes que presentan un nivel de utilización *permanente*, en el grupo 2 (clúster 1) con un 40% nivel *moderado*, mientras que para *ética y derechos humanos* marcado en negrita en la tabla 3.12 muestra el 14% de los estudiantes en un nivel *permanente*, en el grupo 2 (clúster 1) un nivel *moderado* y en el grupo 1 (clúster 0) presenta el 26% *moderado*.
- ❖ **Tareas:** En el caso de esta herramienta se tiene en el grupo 1 (clúster 0) para el curso de *lógica de la programación* un 18% de los estudiantes que tienen un nivel *permanente*, en el caso del grupo 2 (clúster 1) con un 40% *moderado*, mientras que para *ética y derechos humanos* el nivel *permanente* es del 26% en el grupo 1 (clúster 0), en el grupo 2 (clúster 1) con un 14% *moderado*.
- ❖ **Mensajería:** Para esta herramienta se tiene en el grupo 1 (clúster 0) para el curso de *lógica de la programación* un 18% de los estudiantes que tienen un nivel *permanente*, en el grupo 2 (clúster 1) un 40% *moderado*, mientras que para *ética y derechos humanos* es del 26% en el grupo 1 (clúster 0) tienen un nivel *permanente*, en el grupo 2 (clúster 1) un 14% *moderado*.
- ❖ **Twitter:** Esta herramienta en el grupo 1 (clúster 0), del curso de *lógica de la programación* con un 18% de los estudiantes que presentan un nivel de utilización

*permanente*, el caso del grupo 2 (clúster 1) un 40% *moderado*, mientras que para *ética y derechos humanos* marcado en negrita presenta el 14% de los estudiantes en un nivel *permanente* en el grupo 2 (clúster 1), para el nivel *moderado* con un 26% en el grupo 1 (clúster 0).

- ❖ **Cuestionario:** Aquí en el grupo 1 (clúster 0) para el curso de *lógica de la programación* un 18% de los estudiantes tienen un nivel *permanente*, en el grupo 2 (clúster 1) con un 40% en un nivel *moderado*, mientras que para *ética y derechos humanos* el grupo 1 (clúster 0) es del 26% en un nivel *permanente*, para el grupo 2 (clúster 1) presenta un 14% en un nivel *moderado*.

Con fundamento en las particularidades mencionadas con anterioridad, se puede concluir que los estudiantes que presentan más interacción con la herramienta (foros) pertenecen al curso de *ética y derechos humanos*, mientras que en los estudiantes del curso de *lógica de la programación* presentan más interacción con las herramientas (foros, recursos, tareas, mensajería, twitter y cuestionario). Denotando claramente a nivel general que de entre los dos cursos, la utilización de las herramientas *permanente* y *moderado*, representan mayor interacción los estudiantes que corresponden al curso *lógica de la programación* (informática).

## **3.7 Minería de datos para determinar estilos de aprendizaje**

### **3.7.1 Introducción**

El aprendizaje depende fuertemente de la influencia del profesor, de sus alcances, del dominio de su materia, del ámbito de su competencia, del método pedagógico, pero también de su estilo muy particular de enseñanza. En este punto, resulta que los estudiantes también tienen sus características propias de aprendizaje, es decir, cada uno tenemos nuestra manera particular de aprender o estilo de aprendizaje.

Estas son las características que (Felder y Silverman, 1988) afirman y nos revelan los distintos modos que las personas tenemos de visualizar, de captar y percibir el mundo físico, de las diferentes habilidades y que ponemos en juego cuando de aprender se trata. Se poseen desde el nacimiento y diferencian a un individuo de otro. Por esta razón podemos identificar a una persona como muy “lineal”, o creativa, o intuitiva, sensorial, con capacidad de abstracción, etc. Cada una de estas características se manifiesta con diversos matices en cada uno de nosotros la forma y eficiencia con que se utilizan en el aprendizaje nos definen individualmente.

De acuerdo con (Felder, 1990) y (Pérez, 1995), se requiere hacer compatibles los estilos de aprendizaje de los estudiantes con los estilos de enseñanza de los profesores. Mientras más cercanos se hallen éstos, mejor será el rendimiento académico.

Por lo antes mencionado se considera necesario determinar las características del perfil de aprendizaje de los estudiantes de la UTPL, con este propósito encuestamos a los estudiantes de la carrera de Informática modalidad “Abierta” que conforman la población y muestra descrito con más detalle en el punto 3.9.

### 3.7.2 Modelo de estilos de aprendizaje

Un modelo de estilo de aprendizaje clasifica a los estudiantes de acuerdo con la forma en cómo ellos reciben y procesan la información (Durán y Costaguta, 2007).

La categoría bipolar de Felder y Silverman en (Aragón y Jiménez, 2009) citan a (González, 1996), donde indica que quienes desarrollaron un modelo que explica los diferentes estilos de aprendizaje de los alumnos fueron, Richard M. Felder y Linda K. Silverman basados en los siguientes postulados:

- ❖ El tipo de información que recibe el alumno es predominantemente sensitiva o bien intuitiva.
- ❖ La modalidad sensorial utilizada preferentemente por los alumnos es auditiva o visual.
- ❖ Los alumnos tienen dos formas de organizar la información: inductivamente o bien deductivamente.
- ❖ Los alumnos procesan y comprenden la información de dos maneras: secuencialmente o bien globalmente.
- ❖ Los alumnos trabajan con la información recibida de dos maneras: activamente o reflexivamente.

(Aragón y Jiménez, 2009). Estos postulados originan a cinco estilos de aprendizaje, los cuales indican distintas clases entre sí, por lo que se le conoce como modelo de la categoría bipolar. Los estilos de aprendizaje y las características de cada uno son:

### 1. Sensitivos o intuitivos

**Sensitivos:** Resuelven problemas siguiendo procedimientos establecidos, son prácticos y concretos, memorizan con facilidad los hechos, esta clase de estilo de aprendizaje, va dirigido a estudiantes capaces de experimentar y captar mejor la información a través de la sensación que les produce al trabajar por ejemplo con una pintura interpretan y aprenden lo que esta les transmite.

**Intuitivos:** Les gusta innovar; orientados hacia las teorías y los significados; eligen descubrir posibilidades y relaciones; se les facilita el comprender de manera rápida nuevos conceptos; trabajan bien con abstracciones y formulas matemáticas, no les agrada los cálculos rutinarios o mucha memorización en los cursos.

### 2. Auditivos o visuales

**Visuales:** Obtienen la información mediante representaciones visuales como los diagramas de flujo, mapas conceptuales, mapas mentales, se les facilita recordar mejor lo que ven.

**Auditivos:** Prefieren obtener la información en forma hablada, recuerdan mejor lo que escuchan, mediante grabaciones de audio sobre alguna temática de su interés.

### 3. Inductivos o deductivos

**Inductivos:** Comprenden mejor la información cuando se les presentan hechos y observaciones para luego hacer conclusiones hacia los principios o generalizaciones.

**Deductivos:** Prefieren que a partir los principios o las generalizaciones de los hechos que se les presente, deducir ellos mismos las consecuencias y aplicaciones particulares.

#### 4. Secuenciales o globales

**Secuenciales:** Aprenden en pequeños pasos incrementales, su razonamiento es ordenado y lineal; cada paso tiene que estar relacionado con el anterior; solucionando problemas por pasos lógicos.

**Globales:** Aprenden a grandes saltos, con material que obtienen repentinamente visualizan la totalidad; habitualmente resuelven problemas complejos rápidamente de forma innovadora. Sin embargo pueden tener dificultad para explicar como lo hicieron.

#### 5. Activos o reflexivos

**Activos:** Retienen y comprenden mejor nueva información cuando la manejan haciendo algo dinámico con ella (la discuten, la aplican, y explican a los demás). Prefieren trabajar en grupo y ensayar las cosas.

**Reflexivos:** Tienen a retener y comprender nueva información examinando y reflexionando sobre ella; prefieren trabajar solos meditando sobre un tema en específico.

### 3.8 Población y muestra de estudio

La Población de estudio fue conformada por 46 estudiantes que fueron los que dieron contestación al test, y que corresponden a los últimos ciclos de la carrera de “Informática” comprendiendo las asignaturas de: Seminario II(Desarrollo de páginas web dinámicas con php) e Inteligencia artificial, de la UTPL modalidad Abierta, se les realizó el test de Felder ver en anexo C, este test fue habilitado en el EVA a manera de enlace y posteriormente los resultados obtenidos se fueron almacenado en una base de datos, para proceder a realizar luego lo que es la extracción de conocimiento. La variable de estudio es *estilos de aprendizaje*.

*Tabla 3.28: Variable estilos de aprendizaje. Bases Teóricas (Felder y Silverman)*

| <b>Dimensión</b>   | <b>Definición</b>  | <b>Indicador</b>   |
|--|--|--|
| Dimensión relativa a la forma de trabajar con la información:<br><b>Activo – Reflexivo</b>               | La información se puede procesar mediante tareas activas a través de compromisos en actividades físicas o discusiones o a través de la reflexión o introspección.  | <i>¿Cómo prefiere el estudiante procesar la información?</i>   |
| Dimensión relativa al tipo de información:<br><b>Sensorial - Intuitivos</b>                              | Básicamente, los estudiantes perciben dos tipos de información: información externa o sensitiva a la vista, al oído o a las sensaciones físicas e información interna o memorias, ideas, lecturas, etc.  | <i>¿Qué tipo de información perciben preferentemente los estudiantes?</i>                            |
| Dimensión relativa al tipo de estímulos preferenciales:<br><b>Visual - Verbal</b>                        | Con respecto a la información externa, los estudiantes básicamente la reciben en formatos visuales mediante cuadros, diagramas, gráficos, demostraciones, etc. o en formatos verbales mediante sonidos, expresión oral y escrita, fórmulas, símbolos, etc. | <i>¿A través de qué modalidad sensorial es más efectivamente percibida la información cognitiva?</i> |
| Dimensión relativa a la forma de procesar y comprensión de la información:<br><b>Secuencial - Global</b> | El progreso de los estudiantes sobre el aprendizaje implica un procedimiento secuencial que necesita progresión lógica de pasos incrementales pequeños o entendimiento global que requiere de una visión integral.   | <i>¿Cómo progresa el estudiante en su aprendizaje?</i>   |

Con la información obtenida a través del test se generó una base de datos, que fue sometida al proceso de KDD<sup>11</sup>, siguiendo sus tres etapas: pre-procesamiento, minería de datos, y post-procesamiento. La información recabada con el test fue cargada en una planilla de Excel, para su posterior carga a la base de datos minería acciones figura D.2, requerida como entrada para el software WEKA figura D.3.

<sup>11</sup> KDD, Knowledge Discovery in Databases, o Proceso de descubrimiento de conocimiento en bases de datos relacionales.

Entre las actividades de pre procesamiento se eliminaron algunas inconsistencias a través de correcciones manuales, tales como la existencia de espaciados innecesarios. Dado que se trabajó con un único archivo no fue necesario realizar tareas de integración ni de transformación. Se llevó a cabo una única tarea de reducción de datos que consistió en eliminar los atributos: nombre del curso, por resultar irrelevante para los objetivos de este proceso, de descubrimiento de conocimiento los estudiantes como las respuestas a P8 y P16 ya que éstas corresponden al procesamiento de la información de los estudiantes, más no a estilos de aprendizaje.

En la etapa de minería de datos se utilizó el clúster que posee el software WEKA para identificar subgrupos homogéneos dentro de la población de estudiantes a los cuales se les realizó el test de Felder y Silverman, seleccionando el algoritmo *FarthestFirst*, debido a que existen algunas investigaciones como la de (Durán y Costaguta, 2007), en las que ya se ha experimentado con este algoritmo demostrando ser el idóneo para someter los datos a un proceso de KDD e interpretar los resultados obtenidos, pues en esa investigación para detectar los estilos de aprendizaje dominantes optan por mantener en dos la cantidad de él clúster a generar de igual manera se lo hará para el desarrollo de este trabajo .

Tabla 3.29: Correspondencia entre los estilos de aprendizaje. Bases Teóricas (Felder y Silverman)

| ESTILOS DE APRENDIZAJE | CODIGO    | RESPUESTAS CONSIDERADAS   |
|------------------------|-----------|---|
| <b>SENSORIAL</b>       | <b>SN</b> | P2=a, P6=a, P10=a, P14=a, P18=a, P22=a, P26=a, P29=a, P30=a, P34=a, P38=a, P42=a, P43=a |
| <b>INTUITIVO</b>       | <b>I</b>  | P2=b, P6=b, P10=b, P14=b, P18=b, P22=b, P26=b, P29=b, P30=b, P34=b, P38=b, P42=b, P43=b |
| <b>VISUAL</b>          | <b>V</b>  | P3=a, P7=a, P11=a, P15=a, P19=a, P23=a, P27=a, P31=a, P35=a, P39=a                      |
| <b>AUDITIVO</b>        | <b>AU</b> | P3=b, P7=b, P11=b, P15=b, P19=b, P23=b, P27=b, P31=b, P35=b,                            |

|                   |           |  |
|-------------------|-----------|--|
|                   |           | P39=b  |
| <b>ACTIVO</b>     | <b>AC</b> | P1=a, P5=a, P9=a, P13=a, P14=a, P17=a, P21=a, P25=a, P33=a, P37=a, P41=a |
| <b>REFLEXIVO</b>  | <b>R</b>  | P1=b, P5=b, P9=b, P13=b, P14=b, P17=b, P21=b, P25=b, P33=b, P37=b, P41=b |
| <b>SECUENCIAL</b> | <b>SC</b> | P4=a, P12=a, P20=a, P24=a, P28=a, P32=a, P36=a, P40=a, P44=a             |
| <b>GLOBAL</b>     | <b>G</b>  | P4=b, P12=b, P20=b, P24=b, P28=b, P32=b, P36=b, P40=b, P44=b             |

### 3.9 Experimentación y resultados

Ahora se experimentara con el algoritmo **FarthestFirst** que ya se había mencionado con anterioridad que fue el seleccionado teniendo en cuenta que es necesario realizar varios experimentos con este para obtener el resultado más óptimo posible en cuanto a la distribución de las instancias en cada clúster, hacemos uso de este ya que lo que interesa es agrupar los estudiantes homogéneos, además de que se trata de un problema de k centros, donde cada instancia será asignada a un clúster cuyo centro sea el más cercano.

Para el **primer experimento**:

- Desde la aplicación WEKA, seleccionamos el algoritmo clúster FarthestFirst y le configuramos el valor de 2 clústeres ya que contamos con una muestra de población de 46 , para obtener grupos definidos de estudiantes de acuerdo a los estilos de aprendizaje, teniéndolos en los siguientes tipos de estilos de aprendizaje según las bases teóricas de Felder y Silverman: sensorial(SN), intuitivo(I), visual(V), auditivo(AU), activo(AC), reflexivo(R), secuencial(SC) y global(G), dejando los demás valores de configuración por defecto, así como el modo de evaluación del clúster le dejamos en *Use training set* (Morate) en esta opción WEKA entrenará el método con todos los datos disponibles y luego lo aplicara otra vez sobre los mismos. En la tabla

3.30 se puede observar los centroides de cada clúster, también las instancias que corresponden a cada clúster.

*Tabla 3.30: Primer experimento con el algoritmo FarthestFirst – semilla 5*

| Algoritmo     | Clúster   | Instancias clasificadas |
|---------------|---|-------------------------|
| FarthestFirst | 0   | 43 - 93%                |
|               | 1   | 3 - 7%                  |
|               | <pre> FarthestFirst =====  Cluster centroids:  Cluster 0   a a a b b a a a a a b a a b a a b b a a b a a b a b a a a a a b a a a b a a a Cluster 1   a a a a a a a a b a b a b a b b a b a a b b a b a b a b b b a a a a a b a b b  Clustered Instances  0      43 ( 93%) 1       3 (  7%)           </pre> |                         |

Se aprecia que la agrupación de instancias en cada clúster al enfocarse en los porcentajes no son los ideales para, que este experimento sea sujeto al análisis ya que, se necesita que exista una cantidad de agrupamiento por lo menos similares o cercanos entre los clúster, pues como se puede apreciar el clúster 0, es el que envuelve la mayor cantidad de instancias con un valor de 43(93%), en comparación con el clúster 1 que es apenas de 3 (7%) por ello es necesario realizar un segundo experimento.

Para el **segundo experimento**:

- Se cambia únicamente el valor de la semilla a 20 porque este valor se utiliza en la generación de un número aleatorio que es, a su vez, utilizado para hacer la asignación inicial de instancias de clusters, ya que se debe tener en cuenta que, en general los algoritmos de agrupación son muy sencibles a cómo las agrupaciones se asignan inicialmente. Por lo tanto, muchas veces es necesario probar diferentes valores y evaluar los resultados, el resto valores se le deja igual manteniendo el mismo número de clúster que el experimento anterior. En la tabla 3.31 se puede apreciar los resultados obtenidos esta vez.

*Tabla 3.31: Segundo experimento con el algoritmo FarthestFirst – semilla 20*

| Algoritmo     | Clúster   | Instancias clasificadas |
|---------------|---|-------------------------|
| FarthestFirst | 0   | 39 – 85%                |
|               | 1   | 7 – 15%                 |
|               | <pre> FarthestFirst ===== Cluster centroids: Cluster 0   a b a b b a a a a a a a b a b a a a b b a a b b a b b b a b b b a a a a b b b a a a Cluster 1   a a a a a a a a a b a b a b a b b a b a b a a b b a b a b a b b b a a a a a b a b b Clustered Instances 0      39 ( 85%) 1       7 ( 15%)           </pre> |                         |

Aquí se nota que el clúster 0 es el que tiene una mayor agrupación de instancias en relación al clúster 1, le lleva la delantera con un porcentaje considerablemente, por esto se procede a hacer un nuevo experimento hasta encontrar una distribución adecuada.

Para el **tercer experimento**:

- Esta vez se cambia el valor de la semilla a 100, las demás configuraciones se conserva las mismas en cuanto al número de clúster y el modo de evaluación del clúster. En la tabla 3.32 podemos apreciar los resultados obtenidos esta vez.

Tabla 3.32: Tercer experimento con el algoritmo FarthestFirst – semilla 100

| Algoritmo            | Clúster   | Instancias clasificadas |
|----------------------|---|-------------------------|
| <b>FarthestFirst</b> | 0   | 22 – 48%                |
|                      | 1   | 24 – 52%                |
|                      | <pre> FarthestFirst =====  Cluster centroids:  Cluster 0   a a a b b a a b a a a b a b a a a b b b a b b b b b a b a a b a b a a a a b b a b a Cluster 1   a b a a a a a a a b a a b b b b a a a b b a b a a b a a b a a b b b a a b a a a a b  Clustered Instances  0      22 ( 48%) 1      24 ( 52%)           </pre> |                         |

Ahora sí se cuenta con una distribución adecuada de las instancias en cada clúster, y que merece ser sometida a la interpretación y análisis de los resultados para determinar la tendencia de los estudiantes en cuanto al tipo de estilo de aprendizaje que poseen.

Según la tabla 3.29, teniendo como base el código del estilo de aprendizaje se optó por reemplazar cada uno de los atributos incluidos en los dos centroides de los clústers que resultaron por la aplicación del algoritmo, se elaboro tablas en Excel para facilitar el conteo de la concurrencia y correspondencia de cada *estilo de aprendizaje* según las respuestas al test de esta manera los clústers quedaron como se observa en la figura 3.26 y figura 3.27:

**GRUPO 1 – CLÚSTER 0**

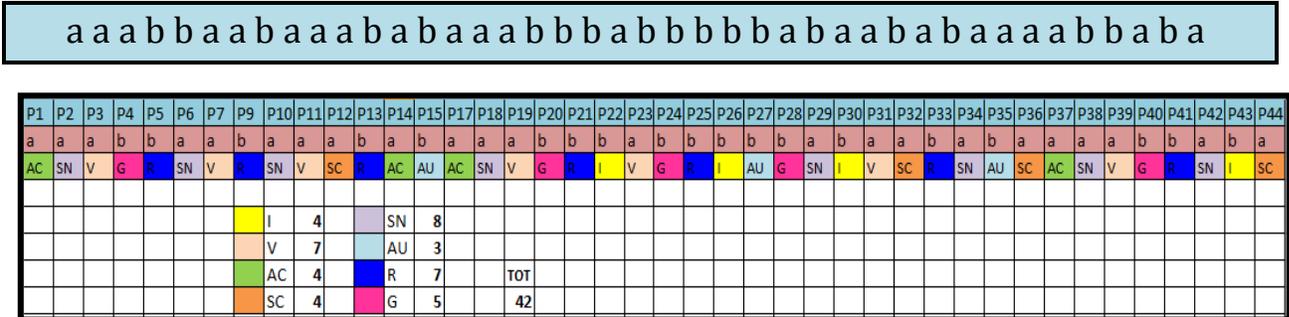


Figura 3.26 Correspondencia del clúster 0

En la figura 3.26 se observa el reemplazo de cada uno de los atributos con su correspondiente estilo de aprendizaje y con la finalidad de establecer la *combinación de estilos de aprendizaje* representada por cada centroide se derivó a contar la cantidad de veces que se repiten de cada estilo teniendo como resultado lo siguiente:

Tabla 3.33: Resultados de correspondencia– clúster 0

| Grupo 1 - Clúster 0   |              |
|-----------------------|--------------|
| Estilo de Aprendizaje | Concurrencia |
| I (Intuitivo)         | 4            |
| V (Visual)            | 7            |
| AC (Activo)           | 4            |
| SC (Secuencial)       | 4            |
| SN (Sensorial)        | 8            |
| AU (Auditivo)         | 3            |
| R (Reflexivo)         | 7            |
| G (Global)            | 5            |
| Total = 42            |              |

Cabe indicar que el total de 42 representa a las preguntas del test, es decir son los atributos, y ¿Por qué son 42? , si en el test de Felder y Silverman son 44 las que se efectúan, pues anteriormente ya se mencionó que se eliminarían los atributos 8 y 16 ya que éstas corresponden a como el estudiante procesa la información más no, a un estilo para aprender, aclarado esto. Continuamos indicando que en el grupo 1 (clúster 0),

según los resultados de la tabla 3.33 se observa marcado en Negrita cuál, de los estilos tuvo mayor concurrencia, reflejando ser el estilo *visual-sensorial* (V-SN), *reflexivo-global* (R-G) que resulta ser el dominante en este grupo.

**GRUPO 2 – CLÚSTER 1**

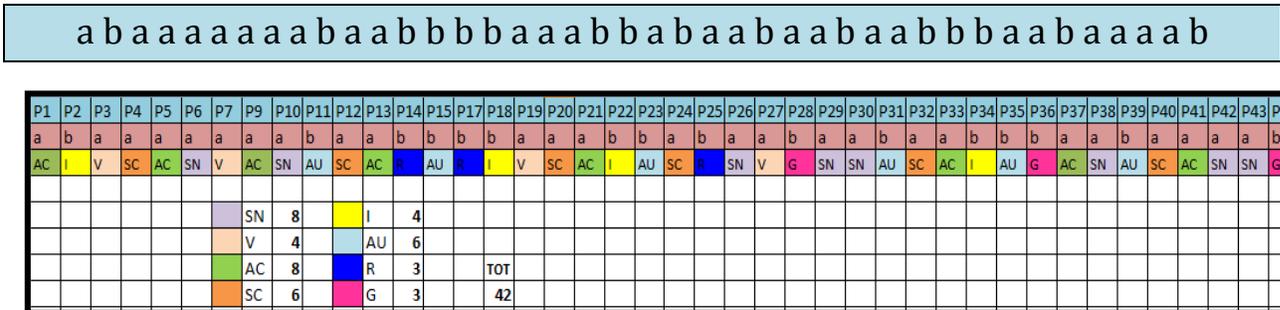


Figura 3.27 Correspondencia del clúster 1

En la figura 3.27 se observa tanto el reemplazo de cada uno de los atributos con su correspondiente estilo de aprendizaje, como la concurrencia de los mismos según las instancias, esto en lo que corresponde al grupo 2 (clúster 1) para ello se procedió a contar la cantidad de veces que se repiten de cada estilo teniendo como resultado lo siguiente:

Tabla 3.34: Resultados de correspondencia– clúster 1

| Grupo 2 – Clúster 1   |              |
|-----------------------|--------------|
| Estilo de Aprendizaje | Concurrencia |
| I (Intuitivo)         | 4            |
| V (Visual)            | 4            |
| AC (Activo)           | <b>8</b>     |
| SC (Secuencial)       | <b>6</b>     |
| SN (Sensorial)        | <b>8</b>     |
| AU (Auditivo)         | <b>6</b>     |
| R (Reflexivo)         | 3            |
| G (Global)            | 3            |
|                       | Total = 42   |

En éstos resultados de la tabla 3.34 que corresponden al grupo 2 (clúster 1), se puede apreciar que, representa al estilo *activo-secuencial* (AC-SC), *sensorial-auditivo* (SN-AU), como consecuencia evidente, que resulta ser dominante en este grupo.

Conclusión de la experimentación con el algoritmo de clustering, los **grupos** que se encontraron al aplicar el algoritmo ***FarthestFirst*** son:

- ❖ **Grupo 1** – Clúster 0: Aquí se encuentran los estudiantes que poseen una mayor relación de concurrencia con el estilo de aprendizaje *visual-sensorial* (V-SN), y *reflexivo-global* (R-G) enunciando, que los primeros optan por obtener y captar mejor la información, con el uso de material visual videos, fotografías, diagramas de flujo, mapas conceptuales, ya que se les facilita el memorizar y recordar mejor lo que observan, mostrando un especial agrado por el trabajo práctico que implica seguir procedimientos fijándose en cada detalle, para luego hacer vínculos y comprender el todo. El *reflexivo-global* (R-G), expresa que los estudiantes tienden a comprender mejor la información cuando ellos trabajan solos, sacan sus conclusiones analizando los hechos sobre algún tema, examinando y reflexionando para asentir lo que han encontrado, visualizando la totalidad de un problema complejo resolviéndolo de manera rápida ejecutando sus propios mecanismos e incluso inventado nuevas formas de resolución de un problema.

Se puede concluir que es un grupo conformado por el 48%, de la población encuestada que manifiesta mayor tendencia para aprender en el estilo **visual-sensorial (V-SN)**, **reflexivo-global (R-G)**, denotando que son estudiantes que aprenden mejor trabajando individualmente, observando siendo cuidadosos y creativos en la forma que realizan sus tareas familiarizándose mejor con imágenes y gráficos para entender todo sobre un tema.

- ❖ **Grupo 2** – Clúster 1: Este grupo está formado por los estudiantes que presentan un mayor índice de asimilación en el estilo de aprendizaje *activo-secuencial* (AC-SC), y *sensorial-auditivo* (SN-AU) lo que manifiesta que los primeros receptan y prestan más atención a algún contenido, cuando lo operan realizan experimentos, en pequeños pasos incrementales, deteniéndose a analizar las evidencias de algún tema en específico para luego realizar deducciones hacia lo descubierto en el tema. El *sensorial-auditivo* (SN-AU), alude a que los estudiantes prefieren dar solución a un problema siguiendo procedimientos concretos y se les facilita el memorizar hechos y situaciones reales, absorben mejor la información en forma hablada ya sea mediante exposiciones en tiempo real o conversación con otras personas.

Se puede concluir que se trata de un grupo conformado por el 52%, de la población encuestada, que presenta mayor tendencia para aprender en el estilo **activo-secuencial (AC-SC)**, **sensorial-auditivo (SN-AU)**, reflejando que son estudiantes que tienen mayor facilidad para trabajar en equipo expresando sus inquietudes, discutiéndolas y escuchando las opiniones de los demás, para luego sacar soluciones concretas que le sean de utilidad.

## CAPÍTULO 4

*Representación de indicadores y atributos  
del estudiante en un modelo de datos.*

## 4.1 Introducción

En este capítulo se presenta el proceso para el diseño de un **modelo de datos** que es un conjunto de conceptos que permite, describir los datos, las relaciones que existen entre ellos, y las condiciones que los datos deben cumplir para reflejar correctamente la realidad deseada, que para fines de este capítulo es **representar los indicadores: nivel de participación en el curso, nivel de utilización de las herramientas, estilos de aprendizaje y atributos del estudiante** en un modelo de datos, tomando como apoyo para el proceso de diseño la metodología E/R<sup>12</sup>.

### 4.1.1 Propuesta del modelo del estudiante usando la minería de datos

En base a los comportamientos sobre las **acciones que el estudiante realiza sobre el eva**, se identificaron ciertas características del estudiante que pueden ser representados en un modelo de datos que represente el perfil de interacción del estudiante en el EVA. En base a estas características se pudieron deducir dos indicadores como: la **participación o interés en el curso** y la **utilización de las herramientas**, que tras aplicar técnicas de minería de datos con los resultados obtenidos, se adquiere **patrones de comportamiento de los estudiantes**, los cuales como se mencionó anteriormente sirven para construir y representar en un modelo de estudiante datos que permitan además predecir el comportamiento futuro de los estudiantes según las acciones que realizan al interactuar en el curso y con las herramientas que ofrece el entorno virtual de aprendizaje de la UTPL. Para cada indicador se requiere de ciertos atributos para determinar sus niveles:

- ❖ Escaso (E)
- ❖ Moderado (M)
- ❖ Permanente (P)

---

<sup>12</sup> E/R. Entidad/Relación

La minería de datos también se utilizó para detectar los **estilos de aprendizaje** que predomina en los estudiantes, y por ende determinar cómo aprenden, el cual también es objeto de representación en el modelo del estudiante, ya que es muy útil conocer sobre las características de aprendizaje en los estudiantes para, de acuerdo a esto los profesores adecúen estrategias de enseñanza y ofrezcan contenidos de acuerdo a las preferencias que tiene el estudiante para retener mejor la información y aprender.

## 4.2 Descripción de la metodología a seguir

En (López), el modelo de datos E-R, propuesto por (Chen, 1976), basado en conceptos: entidades, relaciones y atributos, proporciona una herramienta para representar información del mundo real a nivel conceptual, notación gráfica, tiene semántica clara, fácil de entender aún por no especialistas, independiente de cualquier SABD<sup>13</sup>. Los elementos básicos del modelo E/R son los siguientes:

### ❖ Entidades

Una entidad es un elemento con existencia independiente de otros elementos y se la describe usando un conjunto de atributos.

### ❖ Atributos

Éstos son las propiedades particulares que cada entidad posee. Los atributos pueden ser:

- ❖ Atributos simples o atómicos: son atributos no divisibles.
- ❖ Atributos compuestos: son atributos que se pueden dividir en sus componentes, pudiendo formar jerarquías.
- ❖ Atributos monovaluados: son atributos que tienen un solo valor para una entidad en particular.

---

<sup>13</sup> SABD. Sistema administrador de base de datos

- ❖ Atributos multivaluados: son atributos que tienen límites inferior y superior en el número de valores para una entidad.
- ❖ Atributos almacenados
- ❖ Atributos derivados
- ❖ Valores nulos
- ❖ Atributos complejos: son atributos compuestos o multivaluados anidados de una manera arbitraria (lista, conjuntos).

#### ❖ **Relaciones**

Son las asociaciones entre las entidades.

#### ❖ **Cardinalidad**

Expresa el número de entidades que pueden asociarse vía un tipo de relación, para las relaciones binarias la cardinalidad puede ser:

- Uno a uno (1:1)
- Uno a varios (1:n)
- Varios a uno (n:1)
- Varios a varios (n:n)

Con base en éstos conceptos básicos, que sirven para el proceso de diseño del modelo de datos, en el que se establece el modo en que se relacionan, las entidades entre sí y con referencia a los cuestionamientos que se plantea sobre *¿cómo construir un modelo del estudiante?* sección 1.6.1 del estado del arte, ya se puede proceder a describir que elementos se tomarán en cuenta para representar en el modelo de datos.

### 4.3 Proceso del diseño del modelo de datos del estudiante

Según (Gaudioso, 2002), el modelado del usuario es una tarea que implica una **recogida intensiva de datos** y su posterior proceso. Alguno de éstos datos puede observarlos directamente en el sistema, gracias al uso de consultas SQL directamente a la base de datos, mientras que en contraste a esto existen algunos datos que pueden requerir uno o más pasos adicionales para su adquisición.

Para el diseño del modelo de datos de los estudiantes se toman en cuenta los siguientes aspectos:

#### 4.3.1 Definición de las entidades

En la tabla 4.1 se indica de donde surgen las entidades para el modelo de datos del estudiante puesto que éstas provienen al dar respuesta a las preguntas **¿quién está siendo modelado?** y **¿qué está siendo modelado?**, lo que acopla al propósito de este capítulo mencionado en la introducción, el cual es de **representar los indicadores: nivel de participación en el curso, nivel de utilización de las herramientas, estilos de aprendizaje y atributos del estudiante**, quedando de esta manera establecidas las siguientes entidades.

Tabla 4.1: Definición de entidades

| ¿Quién está siendo modelado?   | ¿Qué está siendo modelado?  |
|--|---|
| <ul style="list-style-type: none"> <li>- Usuario</li> <li>- Curso</li> </ul> | <ul style="list-style-type: none"> <li>- Acciones</li> <li>- Datos de participación en el curso.</li> <li>- Datos de utilización de herramientas</li> <li>- Estilos de aprendizaje</li> </ul> |

### 4.3.2 Definición de los atributos

En la tabla 4.2 se detalla los atributos que son utilizados para describir a las entidades, es decir son las características que éstas tienen, que es lo que hace que se diferencian entre ellas.

*Tabla 4.2: Definición de atributos*

| ENTIDAD           | ATRIBUTOS   |
|-------------------|---|
| <b>mdl_user</b>   | <ul style="list-style-type: none"> <li>- user_id</li> <li>- firstname</li> <li>- lastname</li> <li>- idnumber</li> <li>- city</li> <li>- country</li> <li>- username</li> <li>- s_password</li> <li>- e_mail</li> <li>- address</li> <li>- phone</li> </ul> |
| <b>mdl_course</b> | <ul style="list-style-type: none"> <li>- course_id</li> <li>- fullname</li> <li>- shortname</li> <li>- modalidad</li> <li>- centro</li> </ul>   |
| <b>model_user</b> | <ul style="list-style-type: none"> <li>- model_userid</li> <li>- <b>mdl_user_userid</b></li> <li>- <b>mdl_course_courseid</b></li> </ul>  |

|  |   |
|--|---|
| <p><b>participacion_curso</b></p>      | <ul style="list-style-type: none"> <li>- participacion_cursoid</li> <li>- <b>model_user_model_userid</b></li> <li>- num_accesos_curso</li> <li>- tavg_acceso_curso (tiempo promedio de acceso al curso)</li> <li>- num_accesos_tareas</li> <li>- numveces_sube_tarea</li> <li>- num_veces_descarga_recurso</li> <li>- num_veces_revisa_cuestionario</li> <li>- num_intevaldpb (número de intentos, evaluación a distancia primer bimestre)</li> <li>- tie_evaladpb (tiempo, evaluación a distancia primer bimestre)</li> <li>- num_intevaldsb (número de intentos, evaluación a distancia segundo bimestre)</li> <li>- tie_evaladsb (tiempo, evaluación a distancia segundo bimestre)</li> <li>- num_msn_escritos</li> <li>- num_veces_revisa_historial_msn</li> <li>- num_veces_revisa_perfil</li> <li>- num_veces_actualiza_perfil</li> <li>- num_accesos_foro</li> <li>- num_veces_contesta_foro</li> <li>- num_veces_revisa_discusión_foro</li> <li>- num_veces_actualiza_contest_foro</li> <li>- num_veces_agrega_tema_discusión</li> <li>- num_msn_twitter</li> <li>- nivel_particip_curso</li> </ul> |
| <p><b>utilizacion_herramientas</b></p> | <ul style="list-style-type: none"> <li>- herramientas_id</li> <li>- <b>model_user_model_userid</b></li> <li>- view_forum_forum</li> <li>- add_post_forum</li> <li>- view_discussion_forum</li> <li>- update_postforum</li> <li>- add_discussion</li> <li>- view_resource</li> <li>- view_assignment</li> <li>- upload_upload</li> <li>- ms_twitter</li> </ul>   |

|                                       |   |
|---------------------------------------|---|
|                                       | <ul style="list-style-type: none"> <li>- write_message</li> <li>- history_message</li> <li>- view_quiz</li> <li>- num_intevald<b>pb</b> (número de intentos, evaluación a distancia primer bimestre)</li> <li>- tie_evalad<b>pb</b> (tiempo, evaluación a distancia primer bimestre)</li> <li>- num_intevalad<b>sb</b> (número de intentos, evaluación a distancia segundo bimestre)</li> <li>- tie_evalad<b>sb</b> (tiempo, evaluación a distancia segundo bimestre)</li> <li>- nivel_utilizaci_herramien</li> </ul> |
| <b>resultados_estilos_aprendizaje</b> | <ul style="list-style-type: none"> <li>- resultados_estilo_aprendizajeid</li> <li>- <b>model_user_model_userid</b></li> <li>- nombre_estilo_aprendizaje</li> <li>- valor_resultado_estiloa</li> </ul>   |
| <b>acciones</b>                       | <ul style="list-style-type: none"> <li>- accionesid</li> <li>- nombre_accion</li> <li>- descripcion_accion</li> </ul>   |
| <b>acciones_en_objetos</b>            | <ul style="list-style-type: none"> <li>- acciones_en_objetosid</li> <li>- <b>objetos_objetosid</b></li> <li>- <b>acciones_accionesid</b></li> </ul>   |
| <b>objetos</b>                        | <ul style="list-style-type: none"> <li>- objetoid</li> <li>- nombre_objeto</li> <li>- descripcion_objeto</li> </ul>   |

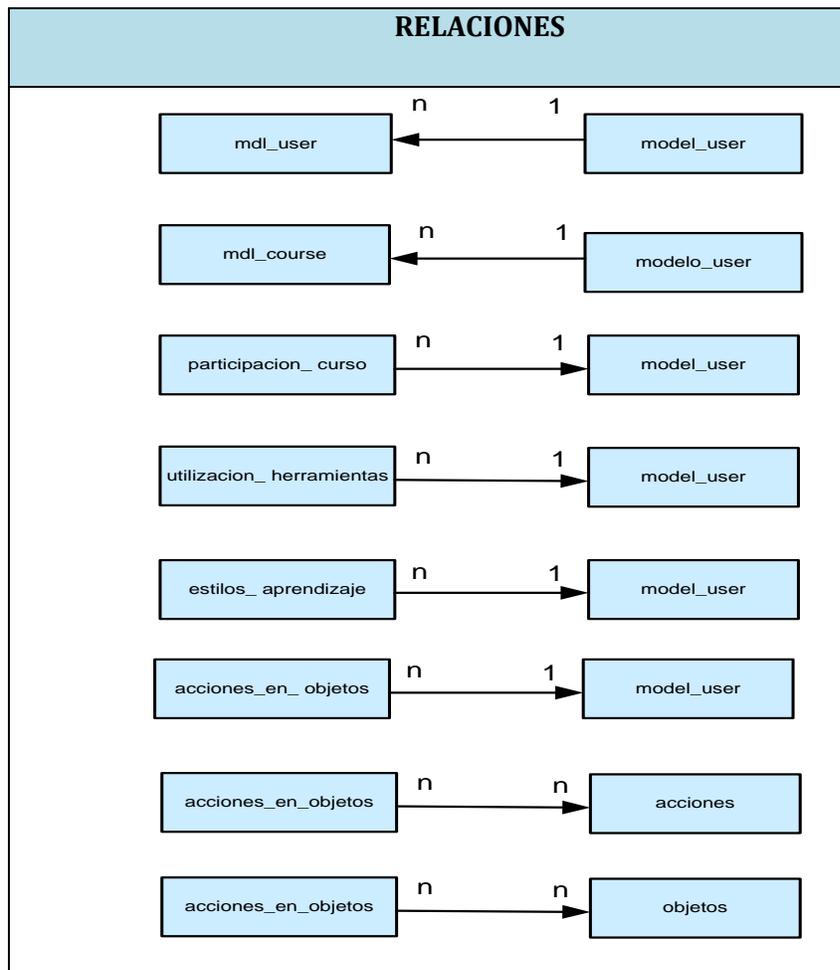
### 4.3.3 Definición de las relaciones y cardinalidad

En (Booch, Rumbaugh y Jacobson, 2004), señalan que una asociación, específica que los objetos de un elemento están conectados con los objetos de otro. Dada una asociación entre dos clases, se puede navegar desde un objeto de una clase hasta un objeto de la otra clase, y viceversa.

Es lógico que ambos extremos de una asociación estén conectados a la misma clase. Eso significa que, dado un objeto de la clase, se puede conectar con otros objetos de la misma clase.

Una asociación que conecta exactamente dos clases se dice binaria, a pesar de que esto no se da muy seguido ya que, se pueden tener asociaciones que conecten más de dos clases; estas son llamadas asociaciones n-arias. Teniendo como fundamento esto, en la tabla 4.3 se definen las relaciones que intervienen en el modelo de datos del estudiante, con su respectiva cardinalidad.

*Tabla 4.3: Definición de relaciones y cardinalidad*



Aquí se puede apreciar las relaciones entre la entidad *modelo\_estudiante*, que es quien se dirige y está asociada con la entidad *estudiante* y la entidad *curso*, posterior a estas relaciones encontramos las de las entidades *participación\_curso*, *utilización\_herramientas*, y *estilos\_aprendizaje* que son quienes se dirigen y se asocian a la entidad *modelo\_estudiante*.

#### 4.4 Modelo de datos del Estudiante

El modelo del estudiante es definido por (VanLehn, 1988), como los componentes de un sistema tutorial inteligente que representan el actual estado de un estudiante.

<sup>14</sup>Los componentes de un modelo individual son:

- ❖ **Creencias conceptuales:** Son las creencias que el alumno posee sobre el problema.
- ❖ **Creencias personales:** Aquellas que el estudiante posee sobre sus compañeros de trabajo.
- ❖ **Acciones individuales:** Las que realiza el individuo en solitario para resolver el problema.
- ❖ **Acciones colectivas:** Se realizan en común, con la finalidad de solucionar el problema.
- ❖ **Objetivos:** La delimitación de objetivos en circunstancias colaborativas es muy importante, ya que existen dos clases de objetivos los individuales y los comunitarios. Se deben lograr que los objetivos individuales no perjudiquen los intereses del grupo ni que los intereses globales del grupo no beneficien a parte del equipo y desfavorezca a otros.
- ❖ **Errores:** La detección de errores es imprescindible para retroalimentar el sistema y proporcionar ayuda oportuna. Cuando se trabaja en grupo la detección de errores es más difícil puesto que una persona puede asumir las ideas de otras sin llegar a creerlas.

El **modelo de datos del estudiante** se presenta en la figura 4.1, donde se realiza ya la unificación de las entidades, atributos y relaciones principales que intervienen, las cuales se han ido describiendo tras el proceso de aplicar la metodología entidad – relación.

---

<sup>14</sup> <http://bit.ly/sI5h2b>

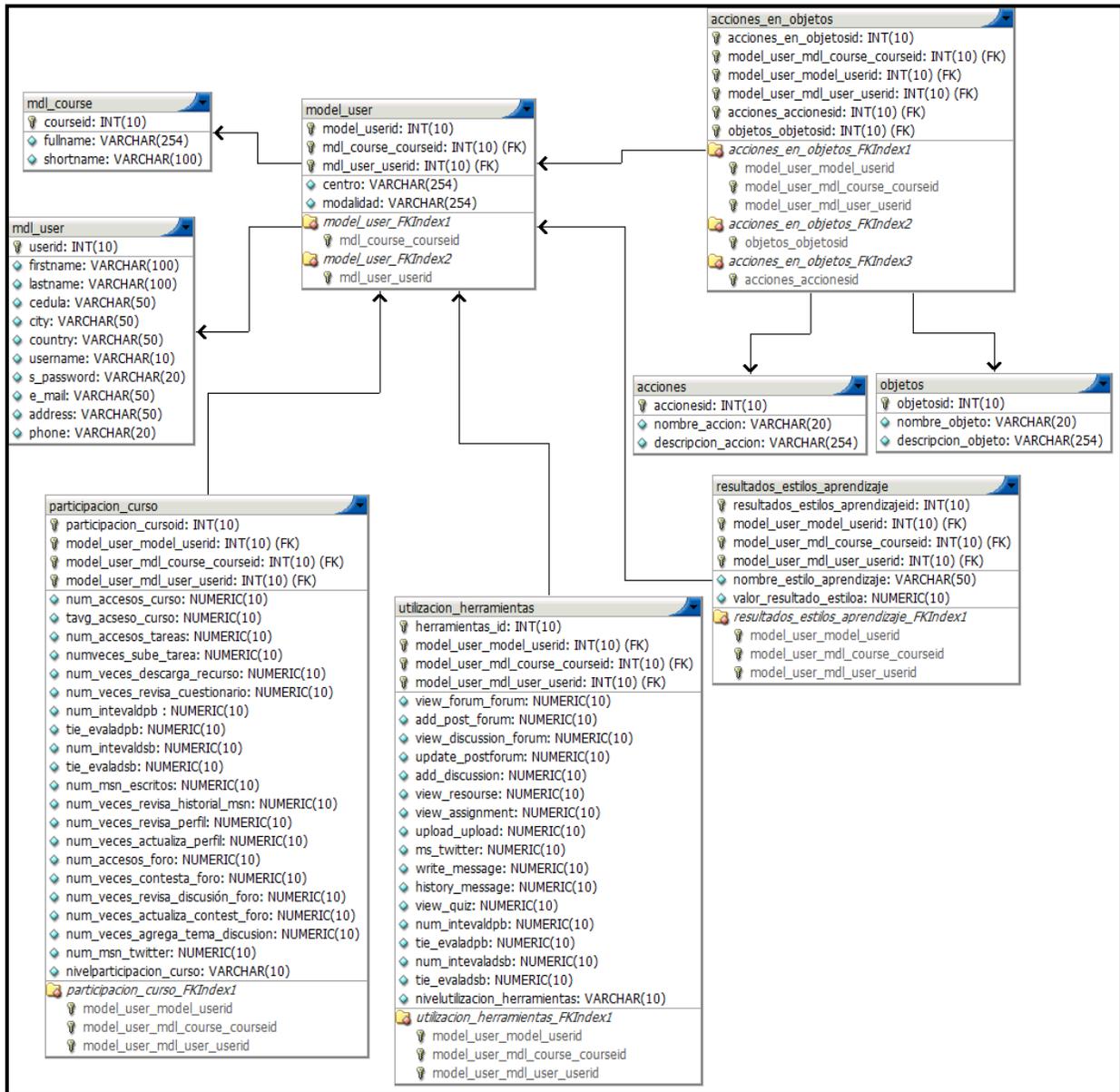


Figura 4.1: Modelo de datos del estudiante

El modelo de datos del estudiante, está conformado por 9 tablas, y está diseñado para el administrador de base de datos SQLyog8.05 de MySQL, cada una es imprescindible debido a la utilidad y función que desempeñan en el sistema, pues las entidades albergan los atributos, así como también expresan las interrelaciones existentes entre sí.

#### 4.4.1 Descripción de las entidades del modelo

Las entidades del modelo de datos, son las destinadas en alojar toda la información relevante del estudiante, expresando la realidad de su nivel de participación en el curso, nivel de utilización de herramientas y el estilo de aprendizaje, dicho en términos de este caso de modelo de datos en particular.

A continuación se describe las entidades que intervienen en este modelo:

- 1) **mdl\_user**: Posee información del estudiante, es la existente en el entorno virtual de aprendizaje, de la cual se recopilan los datos que únicamente interesan, cada atributo que la componen se encuentra descrito en la tabla 4.4

*Tabla 4.4: atributos entidad mdl\_user*

| 1) mdl_user  | DESCRIPCIÓN                                  |
|--------------|--|
| Atributo     |  |
| ✓ user_id    | Identificador del estudiante                 |
| ✓ firstname  | Nombres del estudiante                       |
| ✓ lastname   | Apellidos del estudiante                     |
| ✓ idnumber   | Número de cédula de identidad del estudiante |
| ✓ city       | Ciudad actual del estudiante                 |
| ✓ country    | País actual estudiante                       |
| ✓ username   | Nombre del usuario del estudiante            |
| ✓ s_password | Contraseña del estudiante                    |
| ✓ e_mail     | Correo electrónico del estudiante            |
| ✓ address    | Dirección del estudiante                     |
| ✓ phone      | Número telefónico del estudiante             |

- 2) **mdl\_course**: Tiene información del curso esta entidad, es la propia del entorno virtual de aprendizaje, relevante en este modelo de datos debido a que, para encontrar los indicadores aplicando minería de datos, involucro acceder a esta entidad frecuentemente, ya que de aquí se seleccionaron algunos cursos, que tomó el

estudiante. En la tabla 4.5 se encuentra la descripción de los atributos que únicamente se tomaron en cuenta para la recogida de datos.

*Tabla 4.5: atributos entidad mdl\_course*

| 2) mdl_course | DESCRIPCIÓN                |
|---------------|----------------------------|
| Atributo      |                            |
| ✓ courseid    | Identificador del curso.   |
| ✓ fullname    | Nombre completo del curso. |
| ✓ shortname   | Nombre corto del curso.    |

3) **model\_user**: Esta entidad es propuesta con motivo de asociar los atributos que identifican el modelo de usuario del estudiante, así como también al curso. En la tabla 4.6 se encuentra la descripción de los atributos que intervienen en ella.

*Tabla 4.6: atributos entidad model\_user*

| 3) model_user         | DESCRIPCIÓN  |
|-----------------------|--|
| Atributo              |  |
| ✓ model_userid        | Identificador de la tupla.                                     |
| ✓ mdl_user_user_id    | Indica la relación con la entidad mdl_user.                    |
| ✓ mdl_course_courseid | Indica la relación con la entidad mdl_course.                  |
| ✓ centro              | Nombre del centro universitario en el que se imparte el curso. |
| ✓ modalidad           | Tipo de modalidad de estudios ya sea clásica o a distancia.    |

4) **participacion\_curso**: En esta entidad se guardan los resultados de la participación de los estudiantes en los cursos que fueron seleccionados, para descubrir el nivel de participación. Los atributos que intervienen en esta entidad se encuentran escritos en la tabla 4.7.

*Tabla 4.7: atributos participacion curso*

| 4) participacion_curso             | DESCRIPCIÓN   |
|------------------------------------|---|
| Atributo                           |   |
| ✓ participacion_cursoid            | Identificador de la tupla.  |
| ✓ model_user_model_userid          | Indica la relación con la entidad model_user.   |
| ✓ num_accesos_curso                | Número de veces que el estudiante accede al curso.  |
| ✓ tavg_acceso_curso                | Tiempo promedio de acceso al curso.   |
| ✓ num_accesos_tareas               | Número de veces que accede o revisa tareas.   |
| ✓ numveces_sube_tarea              | Número de veces que envía o sube una tarea.   |
| ✓ num_veces_descarga_recurso       | Número de veces que descarga un recurso.  |
| ✓ num_veces_revisa_cuestionario    | Número de veces que revisa el cuestionario.   |
| ✓ num_intevaldpb                   | Número de intentos en la evaluación a distancia del primer bimestre.                          |
| ✓ tie_evaladpb                     | Tiempo, evaluación a distancia primer bimestre.   |
| ✓ num_intevaldps                   | Número de intentos en la evaluación a distancia del segundo bimestre.                         |
| ✓ tie_evaladsb                     | Tiempo, evaluación a distancia segundo bimestre.  |
| ✓ num_msn_escritos                 | Número de mensajes que escribe el estudiante.   |
| ✓ num_veces_revisa_historial_msn   | Número de veces que revisa el historial de mensajes.  |
| ✓ num_veces_revisa_perfil          | Número de veces que revisa el perfil.   |
| ✓ num_veces_actualiza_perfil       | Número de veces que actualiza el perfil.  |
| ✓ num_accesos_foro                 | Número de accesos al foro.  |
| ✓ num_veces_contesta_foro          | Número de veces que contesta un foro.   |
| ✓ num_veces_revisa_discusion_foro  | Número de veces que revisa una discusión dentro de un foro.                                   |
| ✓ num_veces_actualiza_contest_foro | Número de veces que actualiza contestación de un foro.  |
| ✓ num_veces_agrega_tema_discusion  | Número de veces que agrega un tema de discusión.  |
| ✓ num_msn_twitter                  | Número de mensajes twitter que envía el estudiante.   |
| ✓ nivelparticipacion_curso         | Atributo en el que se guarda el nivel de participación en el curso, por parte del estudiante. |

5) **utilizacion\_herramientas:** En esta entidad se guardan los resultados de la utilización de las herramientas por parte de los estudiantes en los cursos que fueron seleccionados,

para descubrir el nivel de utilización de las herramientas. Los atributos que intervienen en esta entidad se encuentran descritos en la tabla 4.8.

*Tabla 4.8: atributos entidad utilizacion herramientas*

| 5) utilizacion_herramientas     | DESCRIPCIÓN   |
|---------------------------------|---|
| Atributo                        |   |
| ✓ herramientas_id               | Identificador de la tupla.  |
| ✓ model_user_model_userid       | Indica la relación con la entidad model_user.   |
| ✓ view_forum_forum              | Número de accesos al foro.  |
| ✓ add_post_forum                | Número de veces que contesta un foro.   |
| ✓ view_discussion_forum         | Número de veces que revisa una discusión dentro de un foro.                                     |
| ✓ update_post_forum             | Número de veces que actualiza contestación de un foro.  |
| ✓ add_discussion                | Número de veces que agrega un tema de discusión.  |
| ✓ view_resource                 | Número de veces que descarga un recurso.  |
| ✓ view_assignment               | Número de veces que accede o revisa tareas.   |
| ✓ upload_upload                 | Número de veces que envía o sube una tarea.   |
| ✓ ms_twitter                    | Número de mensajes twitter que envía el estudiante.   |
| ✓ write_message                 | Número de mensajes que escribe el estudiante.   |
| ✓ history_message               | Número de veces que revisa el historial de mensajes.  |
| ✓ view_quiz                     | Número de veces que revisa el cuestionario.   |
| ✓ num_intevaldpb                | Número de intentos en la evaluación a distancia del primer bimestre.                            |
| ✓ tie_evaladpb                  | Tiempo, evaluación a distancia primer bimestre.   |
| ✓ num_intevaldsb                | Número de intentos en la evaluación a distancia del segundo bimestre.                           |
| ✓ tie_evaladpb                  | Tiempo, evaluación a distancia segundo bimestre.  |
| ✓ nivelutilizacion_herramientas | Atributo en el que se guarda el nivel de utilización de herramientas, por parte del estudiante. |

**6) resultados\_estilos\_aprendizaje:** En esta entidad se guardan los resultados del test de los estudiantes en los cursos que fueron seleccionados, para descubrir el tipo de estilos

de aprendizaje que posee el estudiante. Los atributos que intervienen en esta entidad se encuentran descritos en la tabla 4.9.

*Tabla 4.9: atributos entidad resultados estilos aprendizaje*

| 6) resultados_estilos_aprendizaje  | DESCRIPCIÓN   |
|------------------------------------|---|
| <b>Atributo</b>                    |   |
| ✓ resultados_estilos_aprendizajeid | Identificador de la tupla.  |
| ✓ model_user_model_userid          | Indica la relación con la entidad model_user.                                       |
| ✓ nombre_estilo_aprendizaje        | Nombre de la tipo de estilo de aprendizaje.   |
| ✓ valor_resultado_estiloa          | Valor del resultado, que el estudiante obtuvo en el test de estilos de aprendizaje. |

**7) acciones\_en\_objetos:** En esta entidad se guardan las acciones que los estudiantes hacen sobre los recursos del EVA. Los atributos que intervienen en esta entidad se encuentran descritos en la tabla 4.10.

*Tabla 4.10: atributos entidad acciones en objetos*

| 7) acciones_en_objetos    | DESCRIPCIÓN                                   |
|---------------------------|---|
| <b>Atributo</b>           |   |
| ✓ acciones_en_objetos_id  | Identificador de la tupla.                    |
| ✓ objetos_id              | Indica la relación con la entidad objetos.    |
| ✓ model_user_model_userid | Indica la relación con la entidad model_user. |
| ✓ acciones_id             | Indica la relación con la entidad acciones.   |

**8) acciones:** En esta entidad se guardan las acciones que los estudiantes pueden ejecutar en los recursos del entorno virtual de aprendizaje, dichas acciones se pueden apreciar en la tabla 3.1. Los atributos que intervienen en esta entidad se encuentran descritos en la tabla 4.11.

*Tabla 4.11: atributos entidad acciones*

| 8) acciones          | DESCRIPCIÓN   |
|----------------------|---|
| Atributo             |   |
| ✓ acciones_id        | Identificador de la tupla.                                      |
| ✓ nombre_accion      | Nombre de la acción que el estudiante puede ejecutar en el EVA. |
| ✓ descripcion_accion | Descripción de la acción que el estudiante ejecuta.             |

**9) objetos:** En esta entidad se establecen los recursos, de los que disponen profesores y estudiantes en el entorno virtual de aprendizaje. Los atributos que intervienen en esta entidad se encuentran descritos en la tabla 4.12.

*Tabla 4.12: atributos entidad objetos*

| 9) objetos           | DESCRIPCIÓN  |
|----------------------|--|
| Atributo             |  |
| ✓ objetos_id         | Identificador de la tupla.                                       |
| ✓ nombre_objeto      | Indica el nombre del objeto (recurso) del EVA                    |
| ✓ descripcion_objeto | Descripción del objeto del en el entorno virtual de aprendizaje. |

## *Discusión*

En este proyecto se desarrolló el proceso de **extracción de conocimiento** aplicando técnicas de minería de datos, para identificar patrones de comportamientos relacionados a las acciones del estudiante con el EVA de la UTPL, a partir de los datos recogidos del periodo Abril2011/Agosto2011 de la modalidad Abierta y en base a las acciones que los estudiantes realizan sobre la plataforma se establecieron dos indicadores, nivel de participación en el curso y nivel de utilización de las herramientas considerando como muestra cursos de las carreras de informática y abogacía.

Las **técnicas utilizadas** fueron de clasificación y de clustering, la primera técnica fue aplicada a las asignaturas dando como resultado la clasificación de los indicadores, para la segunda técnica se realizó la experimentación con una materia de cada carrera a fin de determinar comportamientos en grupos similares y de conocer como aprenden los estudiantes a través del análisis de los estilos de aprendizaje de ciertos estudiantes.

Para el **análisis del conocimiento** se utilizó WEKA debido a que, es una herramienta disponible libremente, que ofrece una variedad de técnicas de minería de datos y puede correr en cualquier plataforma.

## **Resultados de las técnicas de minería de datos**

En la **clasificación**, se experimentó con los algoritmos J48, REPTree, BayesNet, NaiveBayes y JRip con previo un análisis de los resultados que presentó cada uno de éstos, se seleccionó el algoritmo que ofreció los mejores resultados en cuanto al número de clasificaciones correctas para ser sujeto a análisis, resultando ser la mejor opción el REPTree, en el caso de *informática* manifiesta que un 97.6804% presenta un nivel de *participación en el curso* escaso (E) y el 2.3196% está conformado por los estudiantes que presentan un nivel moderado (M) y permanente (P), para el nivel de *utilización de las herramientas* denota un 97.4227% en el nivel escaso y en el 2.5773% están situados los que presentan un nivel moderado (M) y permanente

(P), para el caso de *abogacía* el algoritmo idóneo según sus resultados fue el J48, aquí en el 99.4962% intervienen los que presentan un nivel de *participación en el curso* escaso (E), mientras que en el 0.5038% están ubicados los que tienen un nivel moderado (M) y permanente (P), todo esto refleja un nivel bajo como el dominante en la *participación o interés en el curso* en los estudiantes de las dos carreras.

Las técnicas de **clustering** fueron utilizadas para el indicador del nivel de *utilización de las herramientas*, se experimentó con el algoritmo SimpleKMeans eligiendo a dos grupos de población el primero lógica de la programación (informática), presentando una mayor incidencia en el nivel permanente y moderado en las herramientas (recursos, tareas, mensajería, twitter y cuestionario), el segundo ética y derechos humanos (abogacía), manifestando tener mayor incidencia solamente en la herramienta foros con un nivel moderado, en las demás herramientas presentan un nivel de utilización escaso.

Con el algoritmo FarthestFirst, se analizó los estilos de aprendizaje para descubrir cómo aprenden los estudiantes de los cursos de Seminario II (Desarrollo de páginas web dinámicas con php) e inteligencia artificial, encontrando dos grupos, en el primero están los que presentan un estilo visual-sensorial, reflexivo-global y en el segundo se ubican los que tienen un estilo activo-secuencial, sensorial auditivo.

## **Modelo de datos del estudiante**

En esta parte se plasma en un modelo E/R, todo el trabajo realizado a lo largo del desarrollo del proyecto y partiendo de la recopilación de los datos, teniendo como finalidad encontrar, los indicadores en base a las acciones que el estudiante realiza sobre la plataforma en cada curso, estilos de aprendizaje y los atributos propios de cada estudiante, se diseñó un **modelo de datos** del usuario, en la herramienta DBDesigner 4.

## *Conclusiones y recomendaciones*

## Conclusiones

A partir de las experimentaciones con los datos recogidos y los resultados obtenidos se puede concluir lo siguiente:

- ❖ Con la *minería de datos*, a partir de la información se descubre y analiza nuevo conocimiento útil, para que un sistema cuente con una representación del estado actual del usuario, que en este caso es la descripción de las acciones del estudiante, representado en un modelo que se obtiene a partir de la personalización y análisis de datos, que son aplicaciones propias del *modelado de usuario*.
- ❖ Para predecir el *nivel de participación en el curso* y el *nivel de utilización de las herramientas*, en informática y abogacía después de la aplicación de algoritmos de clasificación, el que presentó los mejores resultados fue el REPTree en informática y J48 en abogacía, y con fundamento en las matrices de confusión presentadas por éstos algoritmos se pudo apreciar notablemente que en ambas carreras existe mayor cantidad de estudiantes que presentan una escasa, participación en el curso, así como en la utilización de herramientas.
- ❖ Se realizó algunas experimentaciones con *clustering* a los datos de los cursos: *lógica de la programación* (informática), y *ética y derechos humanos* (abogacía), el algoritmo seleccionado fue SimpleKMeans para determinar grupos de estudiantes con comportamientos similares. En relación al indicador del *nivel de utilización de las herramientas*, se encontró que para *lógica de la programación* el uso de las herramientas foros, recursos, tareas, mensajería, twitter y cuestionario, es en un nivel *permanente (alto)* y *moderado (medio)*, mientras que en *ética y derechos humanos* los *estudiantes presentan un nivel bajo* de utilización de las herramientas a excepción de los foros dónde se pudo observar que existe un nivel permanente (alto) de interacción y en algunos casos moderado.

- ❖ Haciendo una comparación de los grupos generados en ambas materias, en los resultados se encontró que los estudiantes del curso de lógica de la programación presentan mayor interacción en la mayoría de las herramientas siendo estas: Foros, recursos, tareas, mensajería, twitter y cuestionario, mientras que los estudiantes de ética y derechos humanos revelan mayor interacción en una sola herramienta la cual es foros, por lo que se deduce que a lo mejor en este curso el profesor habilitó para el estudiante más actividades a desarrollar en esta herramienta, que en las demás, demostrando que la cantidad de utilización de las herramientas depende de las actividades que habilite el profesor en ellas.
- ❖ En la herramienta cuestionario, se detectó un grupo de estudiantes que cuando utilizan una mayor cantidad de tiempo por resolver la evaluación a distancia, el número de intentos que realizan es menor, que puede ser porque leen detenidamente las preguntas para proceder a dar contestación, en lugar de hacer varios intentos en menor tiempo que son causa y que pueden ser origen de inseguridad en la respuestas a ingresar.
- ❖ En la búsqueda de los *estilos de aprendizaje* también se realizaron experimentaciones con clustering, se aplicó el algoritmo FarthestFirst, encontrando dos grupos: en el primer grupo se ubican los estudiantes que presentan mayor tendencia en el estilo visual-sensorial (V-SN), reflexivo-global (R-G), reflejando mayor complacencia para trabajar en sus tareas con la ayuda de gráficos, siendo muy observadores y creativos, aprendiendo mejor trabajando solos, en el segundo grupo están los estudiantes que tienen mayor incidencia en el estilo activo-secuencial (AC-SC), sensorial-auditivo (SN-AU), denotando que receptan mejor la información a través de debates con otras personas, exponiendo diferentes puntos de vista y tienen mayor gusto y predisposición para trabajar en grupo.

## Recomendaciones

Con fundamento en el desarrollo del proyecto de Tesis, y de acuerdo a las conclusiones encontradas, en el transcurso de su realización se exponen las siguientes recomendaciones:

- ❖ Para realizar la manipulación de los datos, primero es necesario definir qué es lo que se pretende encontrar en la *materia prima* (datos), y proceder a realizar una exploración de los datos con los que se cuenta.
- ❖ En la selección de las técnicas de minería de datos (diferentes algoritmos), se debe tomar en cuenta, el *dominio del proyecto* ya que la aplicación y efectividad de éstos dependen de él.
- ❖ Después de obtener los niveles predominantes en los estudiantes acerca de los *indicadores* establecidos, es importante incluir estrategias de incentivo para utilizar las herramientas y mejorar la interacción en la plataforma.
- ❖ Para el estudio de los diferentes *estilos de aprendizaje* y debido a la extensa bibliografía existente referente a esto, es sensato seleccionar el que se acople mejor a las características de la población de estudio, en este caso se utilizó el test de Felder y Silverman, para los estudiantes de Informática.
- ❖ Es recomendable, seguir una metodología que se basa en el modelo de datos más difundido en la actualidad, el modelo entidad-relación, para el diseño del *modelo de datos* del usuario, con el objetivo de describir y manipular los datos de un dominio que se desee almacenar en una base de datos.

## Trabajos futuros

Considerar, para trabajos futuros la incorporación de nuevos indicadores que representen acciones de los estudiantes en otras *herramientas* del EVA que no hayan sido consideradas en este trabajo, que a lo mejor hayan sido identificados en otros proyectos afines:

- ❖ La agregación de las acciones que el estudiante realiza en la herramienta chat, con el objeto de descubrir *indicadores de colaboración*, debido a que es un recurso de comunicación entre los participantes de un curso.
- ❖ Establecer *indicadores de navegabilidad* mediante el seguimiento de la ruta que el estudiante efectúa, desde que accede al entorno virtual de aprendizaje, hasta su salida del EVA.
- ❖ Desarrollar un análisis, si el nivel de interacción del estudiante con las herramientas influye en su rendimiento académico con la finalidad de obtener *indicadores de rendimiento académico*.

## Referencias Bibliográficas

- ✍ Alberca, G., Loja, F., Valdiviezo, P. Jiménez, C., y Agila, M. (2008). *Recolección de datos de interacción de alumnos en una plataforma*, Universidad Técnica Particular de Loja. Loja, Ecuador.
- ✍ Anderson, J., Corbett, A., Koeindenger, K., y Pelletier, R. (1995). *Cognitive tutors: Lessons learned. The Journal of the Learning Sciences*, 4(2), 167-207. En Ruiz, F., Molina, L., *Sistema bayesiano para modelado del alumno*. Departamento de Lenguajes y Ciencias de la Computación.
- ✍ Aragón, M., y Jiménez Y. (2009). Diagnóstico de los estilos de aprendizaje en los estudiantes: Estrategia docente para evaluar la calidad educativa. *CPU-e Revista de Investigación Educativa* 9, [http://www.uv.mx/cpue/num9/inves/completos/aragon\\_estilos\\_aprendizaje.pdf](http://www.uv.mx/cpue/num9/inves/completos/aragon_estilos_aprendizaje.pdf)
- ✍ Ayala, A. (2008). *Un modelo del estudiante basado en mapas cognitivos*. Instituto politécnico nacional. México, D.F.
- ✍ Baker, M. (2000). *The roles of models in Artificial Intelligence and Education, International Journal of Artificial Intelligence in Education*, GRIC-COAST, CNRS & Universite Lumiere Lyon. Mendés, France.
- ✍ Bandler, R. y Grinder, J. (1975). *The structure of magic: a book about language and therapy*. Palo Alto, Calif.: Science and Behavior Books.
- ✍ Booch, G., Rumbaugh, J., y Jacobson, I. (2004). *El lenguaje unificado de modelado (UML)*. Madrid: Pearson Addison Wesley.
- ✍ Bressán, G. (2003). *Almacenes de datos y minería de datos*. Retrieved Marzo 01, 2011, from <http://bit.ly/rhwE6Z>
- ✍ Bretch, B. y Jones, M. (1988). "Student models: the genetic graph approach". *International Journal of Man-Machine Studies* , 28, 483-504.
- ✍ Brusilovsky, P., y Millan, E. (2008). *User Models for Adaptive Hypermedia and Adaptive Educational Systems*. USA.

- ✍ Burton, R. y Brown. J. (1982). *An Investigation of Computer Coaching for Informal*. In D. Sleeman & J. S. Brown (Eds.) *Intelligent Tutoring Systems*. New York Academic Press.
- ✍ Chen, P. P. (1976). The Entity-Relationship Model -Toward a Unified View of Data", *Transactions on Database Systems*, Vol. I, No1, March 1976: 9-36 Systems.
- ✍ Dapozo, G., Porcel, E., Lopez, M., Bogado, V., y Bargiela, R. *Aplicación de minería de datos con una herramienta de software libre en la evaluación del rendimiento académico de los alumnos de la carrera de Sistemas de la FACENA-UNNE* .Departamento de Informática. Facultad de Ciencias Exactas y Naturales y Agrimensura. Argentina.
- ✍ Duda, R., Hart, E., y Stork, D. (2001). "Pattern Classification". *Wiley series in Probabilistic and Statistic* (Second Edition). John Wiley and Sons.
- ✍ Duque, M., y Néstor, D. (2008). *Student Model for Adaptive Systems of Virtual Education*. Medellín, Colombia.
- ✍ Duque, N., y Orozco, M. (2011). Minería de Datos para el Análisis de Datos Metereologicos. In J. H. Melo, *Tendencias en Ingeniería de Software e Inteligencia Artificial* (Vol. 4, pp. 105-114).
- ✍ Durán, E., y Costaguta, R. (2007). Minería de Datos para Descubrir Estilos de. *Revista Iberoamericana de Educación* , 6-10.
- ✍ Espinosa, C., Maza, L. y López, A. (2009). Desarrollando un modelo adaptativo jerárquico basado en preferencias de aprendizaje, para entornos e-learning y b-learning para MOODLE.
- ✍ Espinosa, C., Maza, L. y López, A. . (2009). Desarrollando un modelo adaptativo jerárquico basado en preferencias de aprendizaje, para entornos e-learning y b-learning para MOODLE.
- ✍ Felder, R. (1990). Meet your Students. *Revista Chemical Engineering Education* , 7- 8.
- ✍ Felder, R. y Brent R. (2001). Effective strategies for cooperative learning, *J. Cooperation & Collaboration in College Teaching*, 10(2), 69–75. North Carolina State University.

- ✍ Felder, R. y Silverman. (1988). Learning and Teaching Styles, *Engr. Education*, 78(7), 674–681.
- ✍ Finin, T.W. (1989). GUMS – A general user modelling shell. In: *User Models in Dialog Systems* (A. Kobsa & W. Wahlster, eds.). 411-430.
- ✍ Gardner, H. (2003). *La inteligencia reformulada: las inteligencias múltiples en el siglo XXI*. Paidós Ibérica.
- ✍ Gaudioso, E. (2002). *Contribuciones al Modelado del Usuario en Entornos Adaptativos de Aprendizaje y Colaboración a través de Internet*. Tesis Doctoral. Universidad Complutense de Madrid. Madrid, España.
- ✍ Gonzáles de Felipe, T. Guía de Apoyo para el Uso de Moodle 1.9.4 Usuario Alumno. *EUITIO - Universidad de Oviedo* .
- ✍ González, J. (1996). *Estilos cognitivos y de aprendizaje (2a. ed.)*. Barcelona: Santillana.
- ✍ Hernández, J., Ramírez, M., y Ferri, C. (2008). *Introducción a la minería de datos*. España: Pearson Education.
- ✍ Herrmman. (1981). “*The Creative Brain*”. *Training & Development Journal*.
- ✍ Holt, P., Dubs, S., Jones, M., y Greer, J. (1994). The state of student modelling. In *J. Greer & G. McCalla (Eds.), Student Modelling: The Key to Individualized Instruction* .
- ✍ Honney, P. y Mumford, A. (2006). *The Learning Styles Helper’s Guide*. Peter Honey Publications Ltd. Maidenhead.
- ✍ Horvitz, E., Breese, L., Heckerman, D., Hovel, D., y Melse, K. (1998). The lumiere project: Bayesian user modelling for inferring the goals.
- ✍ Jiménez, L., Marcos, J., Fuentes, R. (2010). *Aprendizaje Interactivo de Estructuras de Datos: de la Especificación Algebraica a la Implementación*, Facultad de Informatica Universidad Complutense de Madrid. Madrid, España.
- ✍ Karypis, G., Han, E., y Kumar, V. (1999). “Chameleon: A hierarchical clustering algorithm using dynamic modeling”. *IEEE Computer*.

- ✍ Kobsa, A., Koenemann, J., y Pohl, W. (1999). *Personalized Hypermedia Presentation Techniques for Improving Online Customer Relationships*. Germany.
- ✍ Kolb, D. (1984). *Experiential learning : experience as the source of learning and development*. Englewood Cliffs: Prentice Hall.
- ✍ Leont'ev, A. N. . (1978). *Activity, Consciousness and Personality*. Englewood Cliffs, NJ: Prentice Hall.
- ✍ Linton, F., y Schaefer, H. (2000). *Recommender Systems for Learning: Building User and Expert Models through Long-Term Observation of Application Use*. En Kluwer Academic Publishers (Eds.). *User Modeling and User-Adapted Interaction* 10.(pp. 181-207).
- ✍ López, A. *El Modelo Entidad-Relación*, Posgrado en Ciencia e Ingeniería de la Computación.
- ✍ Martins, A., Faria, L., Vaz de Carvalho, C., y Carrapatoso, E. . (2008). *User Modeling in Adaptive Hypermedia Educational Systems*. Educational Technology & Society. Portugal.
- ✍ Mitchell y Tom. (1997). *Machine Learning*. Ed. McGraw-Hill.
- ✍ Mladen, W. (2008, Abril). *Data Mining y Data Warehousing*.
- ✍ Morate, D. (n.d.). *Manual de Weka*.
- ✍ Moreno Sabido, M. R. (2002, Mayo). *Modelado del Estudiante para un Ambiente de Aprendizaje de Lecto/Escritura*. Tesis Maestría. Universidad de las Américas. Puebla.
- ✍ Muñoz, E., y Muñoz, J. *Interactividad en ambientes virtuales de aprendizaje: Características*. Centro de Ciencias Básicas, Universidad Autónoma de Aguascalientes.
- ✍ O'Connor, J., y Seymour, J. (1995). *Introducción a la PNL*. The Aquarian Press: Urano.
- ✍ Olmos, J., y Gonzales, A. (2007). *Minería de Datos*. Semana de Informática 2007 tu contacto de actualización, Universidad Politécnica de Puebla. San Mateo Cuanalá, Puebla, México.
- ✍ Paredes, P., y Rodríguez, P. (2004). *A mixed approach to modelling styles in adaptive educational hypermedia*. Advanced Technology for Learning.

- ✍ Pérez, T. (1995). Cuestión de Estilo: Enseñanza - aprendizaje en las carreras de ingeniería. *Revista Semillero, de la UABC* . No. 9 , 40- 4.
- ✍ Pinar, L. (2007). *La minería de datos, el aprendizaje automatico y sus distintas técnicas*. Valencia.
- ✍ Quiroz, J. *El rol moderador del tutor en la conferencia mediada por computadora*. Universidad de Santiago de Chile.
- ✍ Romero, C., Ventura, S., y García, E. (2007). *Data mining in course management systems: Moodle case study and tutorial*. University of Córdoba. Córdoba, Spain.
- ✍ Ros, I. (2008). Moodle, la plataforma para la enseñanza y organización escolar . *Ikastorratza, e- Revista* , [http://www.ehu.es/ikastorratza/2\\_alea/moodle.pdf](http://www.ehu.es/ikastorratza/2_alea/moodle.pdf).
- ✍ Santarosa, L. (1999). Escola Virtual: Ambientes de Aprendizaje alternativos para la educación. *Tecnología v Comunicación educativa*. .
- ✍ Self, J.A. (1992). Formal approaches to student modelling, in G. McCalla and J.E. *Student Modelling* .
- ✍ Stauffer, K. (1996). *Student Modeling and Web-based Learning Systems*.
- ✍ Talavera, L., y Gaudioso E. (2004). *Mining student data to characterize similar behavior groups in unstructured collaboration spaces*. Workshop on artificial intelligence in CSCL. 16th European conference on artificial intelligence.
- ✍ Unne Faccena. (2003). MINERÍA DE DATOS, Teleprocesos y Sistemas Distribuidos.
- ✍ VanLehn, K. (1988). *Student Modeling*. Intelligent Tutoring Systems.
- ✍ Vélez, J. (2009). *Entorno de aprendizaje adaptativo soportado por un modelo de usuario integral*. Universitat de Girona.
- ✍ WebMining Consultores. (2011). *KDD: Proceso de Extracción de conocimiento*, Business Intelligence & Analytics, Data Mining.
- ✍ Woolfolk, A. (1996). *Psicología Educativa*. Psicología Educativa: Prentice Hall.

✍ Zaitseva, L., Boule, C. (2003). *Student Models in Computer-Based Education*.

# ANEXOS

## ANEXO A

### *Migración de la Base de Datos*

## A.1 Migración de la base de datos

Para efectuar la migración de los datos se utilizó la herramienta SQLyog8.05 un administrador de MySQL, que permite manipular los datos, así como generar las consultas que se desee de una manera correcta y rápida.

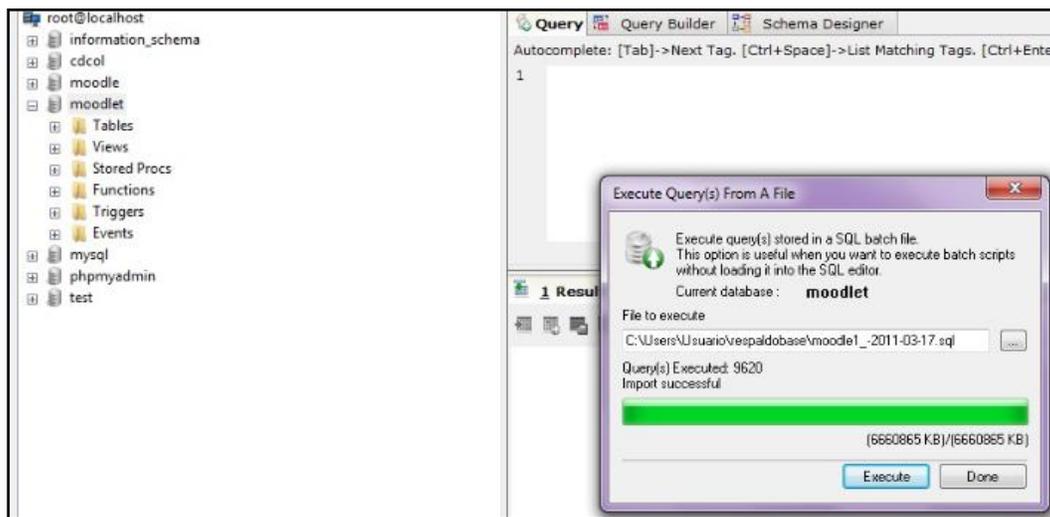


Figura A.1: Migración de DB

Adicional a esto también se realizó la instalación de Moodle1.9.10 alojando así la BD aquí puesto que se puede observar los datos con la interfaz gráfica que presta el EVA.



Figura A.2: Migración de DB a Moodle

## ANEXO B

### *Consultas SQL para Preprocesamiento de los datos.*

## B.1 Consultas SQL

Las consultas que se indican en la tabla B.1 fueron construidas para identificar las asignaturas, que serán tomadas son tomadas como población para el experimento.

### a) Informática

| <b>LOGICA DE LA PROGRAMACIÓN</b> |  |                               |
|----------------------------------|--|-------------------------------|
| 1                                | SELECT DISTINCT  |                               |
| 2                                | mdl_enrol_utpl.courseid,   |                               |
| 3                                | mdl_enrol_utpl.pdoid,  |                               |
| 4                                | mdl_course.fullname  |                               |
| 5                                | FROM   |                               |
| 6                                | mdl_enrol_utpl   |                               |
| 7                                | INNER JOIN mdl_course ON mdl_enrol_utpl.courseid = mdl_course.id |                               |
| 8                                | WHERE  |                               |
| 9                                | mdl_enrol_utpl.pdoid = 207 AND                                   |                               |
| 10                               | mdl_enrol_utpl.courseid = 32885                                  |                               |
|                                  | courseid   | pdoid                         |
|                                  | 32885  | 207                           |
|                                  |  | LOGICA DE LA PROGRAMACION [A] |
| <b>MATEMATICAS DISCRETAS</b>     |  |                               |
| 1                                | SELECT DISTINCT  |                               |
| 2                                | mdl_enrol_utpl.courseid,   |                               |
| 3                                | mdl_enrol_utpl.pdoid,  |                               |
| 4                                | mdl_course.fullname  |                               |
| 5                                | FROM   |                               |
| 6                                | mdl_enrol_utpl   |                               |
| 7                                | INNER JOIN mdl_course ON mdl_enrol_utpl.courseid = mdl_course.id |                               |
| 8                                | WHERE  |                               |
| 9                                | mdl_enrol_utpl.pdoid = 207 AND                                   |                               |
| 10                               | mdl_enrol_utpl.courseid = 33068                                  |                               |
|                                  | courseid   | pdoid                         |
|                                  | 33068  | 207                           |
|                                  |  | MATEMATICAS DISCRETAS [A]     |

**INTELIGENCIA ARTIFICIAL**

```

1  SELECT DISTINCT
2  mdl_enrol_utpl.courseid,
3  mdl_enrol_utpl.pdoid,
4  mdl_course.fullname
5  FROM
6  mdl_enrol_utpl
7  INNER JOIN mdl_course ON mdl_enrol_utpl.courseid = mdl_course.id
8  WHERE
9  mdl_enrol_utpl.pdoid = 207 AND
10 mdl_enrol_utpl.courseid = 32838

```

|                          | courseid | pdoid | fullname                    |
|--------------------------|----------|-------|-----------------------------|
| <input type="checkbox"/> | 32838    | 207   | INTELIGENCIA ARTIFICIAL [A] |

**SEMINARIO II DESARROLLO DE APLICACIONES WEB DINÁMICAS CON PHP**

```

1  SELECT DISTINCT
2  mdl_enrol_utpl.courseid,
3  mdl_enrol_utpl.pdoid,
4  mdl_course.fullname
5  FROM
6  mdl_enrol_utpl
7  INNER JOIN mdl_course ON mdl_enrol_utpl.courseid = mdl_course.id
8  WHERE
9  mdl_enrol_utpl.pdoid = 207 AND
10 mdl_enrol_utpl.courseid = 32860

```

|                          | courseid | pdoid | fullname         |
|--------------------------|----------|-------|------------------|
| <input type="checkbox"/> | 32860    | 207   | SEMINARIO II [A] |

Tabla B.1: Consultas para identificar el curso y su periodo correspondiente (informática).

## b) Abogacía

| <b>MEDICINA LEGAL</b>   |          |       |   |
|---|----------|-------|---|
| <pre> 1  SELECT DISTINCT 2  mdl_enrol_utpl.courseid, 3  mdl_enrol_utpl.pdoid, 4  mdl_course.fullname 5  FROM 6  mdl_enrol_utpl 7  INNER JOIN mdl_course ON mdl_enrol_utpl.courseid = mdl_course.id 8  WHERE 9  mdl_enrol_utpl.pdoid = 207 AND 10 mdl_enrol_utpl.courseid = 33035 </pre> |          |       |   |
| <input type="checkbox"/>  | courseid | pdoid | fullname  |
| <input type="checkbox"/>  | 33035    | 207   | MEDICINA LEGAL [A]                              |
| <b>ETICA Y DERECHOS HUMANOS</b>   |          |       |   |
| <pre> 1  SELECT DISTINCT 2  mdl_enrol_utpl.courseid, 3  mdl_enrol_utpl.pdoid, 4  mdl_course.fullname 5  FROM 6  mdl_enrol_utpl 7  INNER JOIN mdl_course ON mdl_enrol_utpl.courseid = mdl_course.id 8  WHERE 9  mdl_enrol_utpl.pdoid = 207 AND 10 mdl_enrol_utpl.courseid = 33045 </pre> |          |       |   |
| <input type="checkbox"/>  | courseid | pdoid | fullname  |
| <input type="checkbox"/>  | 33045    | 207   | ETICA Y DERECHOS HUMANOS [A]                    |
| <b>LEGISLACION MERCANTIL MONETARIA Y BANCARIA</b>   |          |       |   |
| <pre> 1  SELECT DISTINCT 2  mdl_enrol_utpl.courseid, 3  mdl_enrol_utpl.pdoid, 4  mdl_course.fullname 5  FROM 6  mdl_enrol_utpl 7  INNER JOIN mdl_course ON mdl_enrol_utpl.courseid = mdl_course.id 8  WHERE 9  mdl_enrol_utpl.pdoid = 207 AND 10 mdl_enrol_utpl.courseid = 33501 </pre> |          |       |   |
| <input type="checkbox"/>  | courseid | pdoid | fullname  |
| <input type="checkbox"/>  | 33501    | 207   | LEGISLACION MERCANTIL, MONETARIA Y BANCARIA [B] |

Tabla B.2: Consultas para identificar el curso y su periodo correspondiente (abogacía).

## B.2 Consultas SQL para identificar los registros de las acciones en cada modulo y curso.

En esta consulta figura B.1 se identifican las acciones que el estudiante realiza en el EVA, se creyó conveniente colocar como ejemplo la utilizada con el *courseid* = 32885, el identificador del curso de lógica de la programación, para ejemplo a nivel general ya que las consultas acerca de las acciones en los demás cursos diferirán solo en el valor del atributo del *courseid*.

```
1  SELECT DISTINCT
2  mdl_enrol_utpl.id,
3  COUNT(mdl_log.action),
4  mdl_log.userid,
5  mdl_log.module,
6  mdl_log.action,
7  mdl_log.course,
8  mdl_enrol_utpl.rol
9  FROM
10 mdl_enrol_utpl
11 INNER JOIN mdl_log ON mdl_log.userid = mdl_enrol_utpl.userid
12 WHERE
13 mdl_enrol_utpl.courseid = 32885 AND
14 mdl_log.course = 32885 AND
15 mdl_enrol_utpl.rol <> 3
16 GROUP BY
17 mdl_log.action,
18 mdl_log.userid,
19 mdl_log.module
```

Figura B.1: Consulta para ver todas las acciones que realiza en estudiante en el Eva.

## B.3 Consultas SQL para identificar los registros de los intentos para rendir un cuestionario con sus respectivos tiempos.

En esta consulta figura B.2 se identifican el numero de intentos por resolver un cuestionario, el tiempo de inicio y tiempo de finalización en rendir el cuestionario en el EVA por parte del estudiante, se creyó conveniente colocar como ejemplo la consulta utilizada con el *courseid* = 32885, el identificador del curso de lógica de la programación, para ejemplo a nivel general ya

que las consultas sobre los cuestionarios en los demás cursos diferirán solo en el valor del atributo del *courseid*.

```
1  SELECT DISTINCT
2  mdl_quiz_attempts.userid,
3  mdl_quiz_attempts.timestart,
4  mdl_quiz_attempts.timefinish,
5  mdl_quiz_attempts.quiz,
6  mdl_quiz_attempts.attempt
7  FROM
8  mdl_quiz_attempts
9  INNER JOIN mdl_enrol_utpl ON mdl_enrol_utpl.userid = mdl_quiz_attempts.userid
10 WHERE
11 mdl_enrol_utpl.courseid = 32885 AND
12 mdl_enrol_utpl.rol <> 3
13 -- AND
14 -- mdl_quiz_attempts.quiz = 1320299218
15 AND
16 mdl_quiz_attempts.quiz = 1320300943
17 GROUP BY
18 mdl_quiz_attempts.userid,
19 mdl_quiz_attempts.quiz,
20 mdl_quiz_attempts.timestart,
21 mdl_quiz_attempts.timefinish
```

Figura B.2: Consulta para ver registros de los cuestionarios

## ANEXO C

### *Test de estilos de aprendizaje*

## C.1 Test de Estilos de Aprendizaje

Aquí está detallado el cuestionario del test de estilos de aprendizaje que se realizó a estudiantes de la UTPL, para según resultados evaluar y concluir una mejor manera de enseñanza en el virtual de aprendizaje de la UTPL (EVA).

## C.2 Test de Felder y Silverman

**1. Entiendo mejor algo si:**

- a) Lo práctico
- b) Lo pienso

**2. Me considero:**

- a) Realista
- b) Innovador

**3. Cuando pienso acerca de lo que hice ayer, lo hago gracias a:**

- a) Una imagen
- b) Las palabras

**4. Tengo tendencia a:**

- a) Entender los detalles de un tema pero no ver claramente su estructura completa
- b) Entender la estructura completa pero no ver claramente los detalles

**5. Cuando estoy aprendiendo algo nuevo, me ayuda:**

- a) Hablar acerca de ello
- b) Pensar acerca de ello

**6. Si yo fuera profesor, me gustaría dar un curso:**

- a) Que trate sobre hechos y situaciones reales de la vida
- b) Que trate con ideas y teorías

**7. Prefiero obtener información nueva de:**

- a) Imágenes, diagramas, gráficas o mapas
- b) Instrucciones escritas o información verbal

**8. Cuando entiendo:**

- a) Entiendo todas las partes, entiendo el total
- b) Entiendo el total de algo, entiendo como encajan sus partes

**9. En un grupo de estudio que trabaja con un material difícil, es más probable que:**

- a) Yo participe y contribuya con ideas
- b) Yo no participe y solo escuche

**10. Para mí es más fácil:**

- a) Aprender hechos
  - b) Aprender conceptos
- 11. En un libro con muchas imágenes y gráficas prefiero:**
- a) Revisar cuidadosamente las imágenes y las gráficas
  - b) Concentrarme en el texto escrito
- 12. Cuando resuelvo problemas de matemáticas:**
- a) Habitualmente trabajo sobre las soluciones con un paso a la vez
  - b) Frecuentemente sé cuáles son las soluciones, pero luego tengo dificultad para imaginarme los pasos para llegar a ellas
- 13. En las clases que he tomado:**
- a) He llegado a saber cómo son muchos de los estudiantes
  - b) Raramente he llegado a saber cómo son muchos estudiantes
- 14. Cuando leo temas que no son de ficción, prefiero:**
- a) Algo que me enseñe nuevos hechos o me diga cómo hacer algo
  - b) Algo que me dé nuevas ideas en que pensar
- 15. Me gustan los profesores:**
- a) Que utilizan muchos esquemas en el pizarrón
  - b) Que toman mucho tiempo para explicar algo
- 16. Cuando estoy analizando un cuento o una novela:**
- a) Pienso en los incidentes y trato de acomodarlos para especificar los temas
  - b) Me doy cuenta de cuáles son los temas cuando termino de leer y luego tengo que regresar y encontrar los incidentes que los demuestran
- 17. Cuando comienzo a resolver un problema de tarea, es más probable que:**
- a) Empiezo a trabajar en su solución inmediatamente
  - b) Primero trate de entender completamente el problema
- 18. Prefiero la idea de:**
- a) Certeza
  - b) Teoría
- 19. Recuerdo mejor**
- a) Lo que veo
  - b) Lo que escucho
- 20. Es más importante para mí que un profesor:**
- a) Exponga el material en pasos secuenciales claros
  - b) Me dé una visión general y relacione el material con otros temas
- 21. Prefiero estudiar:**
- a) En grupo de estudio
  - b) Solo
- 22. Me considero:**
- a) Cuidadoso en los detalles de mí trabajo
  - b) Creativo en la forma en la que hago mí trabajo

**23. Cuando alguien me da direcciones de nuevos lugares, prefiero:**

- a) Un mapa
- b) Instrucciones escritas

**24. Aprendo:**

- a) A un paso constante. Si estudio con ahínco consigo lo que deseo
- b) En inicios y pausas. Me llevo a confundir y repentinamente lo entiendo

**25. Prefiero primero**

- a) Hacer algo y ver qué sucede
- b) Pensar cómo voy a hacer algo

**26. Cuando leo por diversión, me gustan los escritores que:**

- a) Dicen claramente lo que desean dar a entender
- b) Dicen las cosas en forma creativa e interesante

**27. Cuando veo un esquema o bosquejo en clase, es más probable que recuerde:**

- a) La imagen
- b) Lo que el profesor dijo acerca de ella

**28. Cuando me enfrento a un conjunto de información:**

- a) Me concentro en los detalles y pierdo de vista el total de la misma.
- b) Trato de entender el todo antes de ir a los detalles.

**29. Recuerdo más fácilmente:**

- a) Algo que he hecho
- b) Algo en lo que he pensado mucho

**30. Cuando tengo que hacer un trabajo, prefiero:**

- a) Dominar una forma de hacerlo
- b) intentar nuevas formas de hacerlo

**31. Cuando alguien me enseña datos, prefiero:**

- a) Gráficas
- b) Resúmenes con texto

**32. Cuando escribo un trabajo, es más probable que:**

- a) Lo haga (pensar o escribir) desde el principio y avanza
- b) Lo haga (pensar o escribir) en diferentes partes y luego las ordene

**33. Cuando tengo que trabajar en un proyecto de grupo, primero quiero:**

- a) Realizar una "lluvia de ideas" donde cada uno contribuye con ideas
- b) Realizar la "lluvia de ideas" en forma personal y luego juntarme con el grupo para comparar las ideas

**34. Considero que es mejor elogio llamar a alguien:**

- a) Sensible
- b) Imaginativo

**35. Cuando conozco gente en una fiesta, es más probable que recuerde:**

- a) Cómo es su apariencia

- b) Lo que dicen de ellos mismos
- 36. Cuando estoy aprendiendo un tema, prefiero:**
  - a) Mantenerme concentrado en ese tema, aprendiendo lo más que pueda de él
  - b) hacer conexiones entre ese tema y temas relacionados
- 37. Me considero:**
  - a) Abierto
  - b) Reservado
- 38. Prefiero cursos que dan más importancia a:**
  - a) Material concreto (hechos, datos)
  - b) Material abstracto (conceptos, teorías)
- 39. Para divertirme, prefiero:**
  - a) Ver televisión
  - b) leer un libro
- 40. Algunos profesores inician sus clases haciendo un bosquejo de lo que enseñarán. Esos bosquejos son:**
  - a) Algo útil para mí
  - b) Muy útiles para mí
- 41. La idea de hacer una tarea en grupo con una sola calificación para todos:**
  - a) Me parece bien
  - b) No me parece bien
- 42. Cuando hago grandes cálculos:**
  - a) Tiendo a repetir todos mis pasos y revisar cuidadosamente mi trabajo
  - b) Me cansa hacer su revisión y tengo que esforzarme para hacerlo
- 43. Tiendo a recordar lugares en los que he estado:**
  - a) Fácilmente y con bastante exactitud.
  - b) Con dificultad y sin mucho detalle.
- 44. Cuando resuelvo problemas en grupo, es más probable que yo:**
  - a) Piense en los pasos para la solución de los problemas.
  - b) Piense en las posibles consecuencias o aplicaciones de la solución en un amplio rango de campos.

### Test de Estilos de Aprendizaje

Seleccione la opción "a" o "b" para indicar su respuesta a cada pregunta. Por favor seleccione solamente una respuesta para cada pregunta. Si tanto "a" y "b" parecen aplicarse a usted, seleccione aquella que se aplique más frecuentemente. No es necesario pensar mucho para dar su respuesta.

**SELECCIONE LA ASIGNATURA QUE CORRESPONDE :**

- a) Inteligencia Artificial
- b) Desarrollo de páginas web dinámicas con PHP

**1. Entiendo mejor algo si:**

- a) Lo práctico
- b) Lo pienso

**2. Me considero:**

- a) Realista
- b) Innovador

**3. Cuando pienso acerca de lo que hice ayer, lo hago gracias a:**

- a) Una imagen
- b) Las palabras

*Figura C.1: Test de estilos de aprendizaje habilitado en cada curso.*

## ANEXO D

### *Carga de los datos en WEKA*

## D.1 Conexión de la base de datos MineríaAcciones en mysql con Weka

Adicional a los archivos .csv se creó una base de datos llamada MineríaAcciones, donde recopilamos todas las acciones, éstos datos son los mismos que se ha venido utilizando para efectos de la minería de datos, aprovechando que Weka permite la posibilidad de conectar directamente a una base. Para esto vamos al boton **Open DB** damos clic y nos aparecerá la ventana figura D.1.

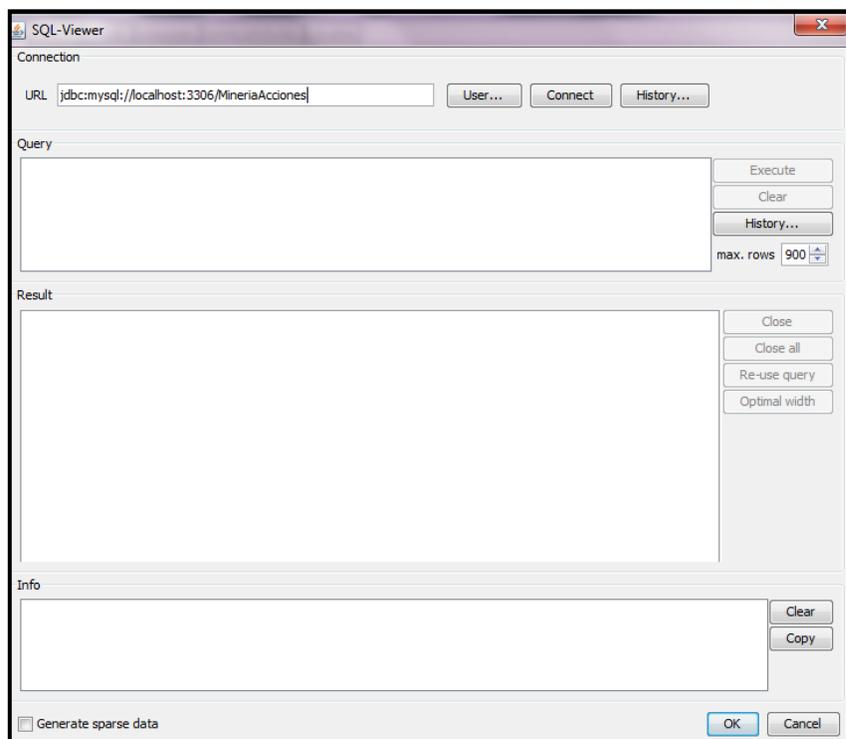


Figura D.1 Conexión a la base en WEKA con mysql.

Seguidamente vamos al botón **user** para configurar los parámetros que se muestran a continuación figura D.2 ya que son indispensables para que la conexión se realice con éxito.

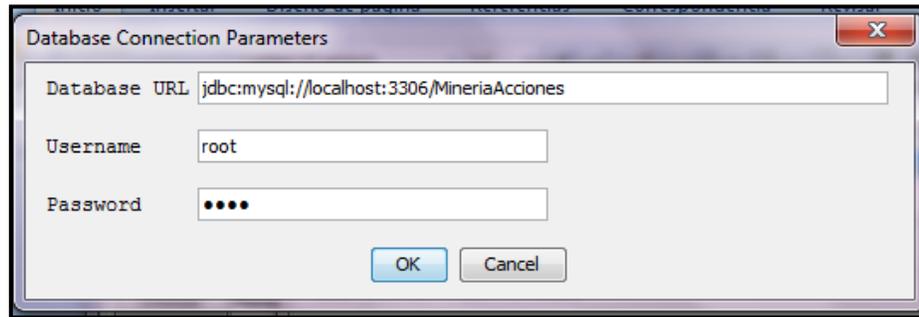


Figura D.2 Parámetros de conexión

Después damos clic en **OK**, luego en **connect**, inmediatamente en el panel de **info** nos notificara si la conexión es TRUE o FALSE, cuando se haya conectado **TRUE** podemos escribir ya la consulta que se desee. Finalmente le damos clic en **Execute** mostrandonos una ventana como la que sigue figura D.3.

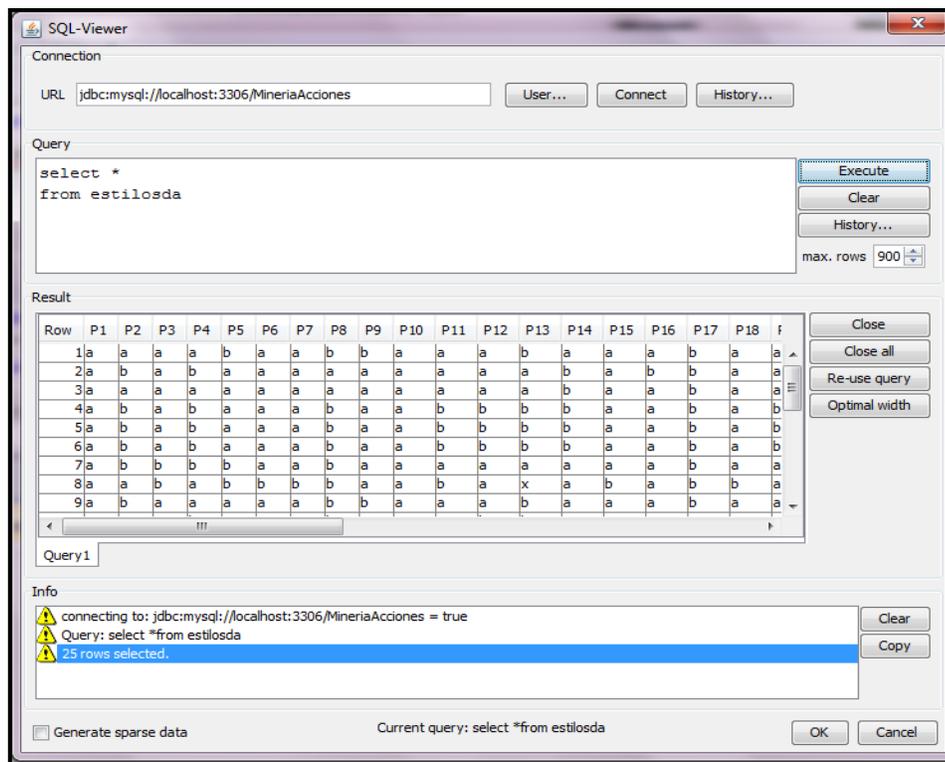


Figura D.3 Conexión a la base MineriaAcciones en Weka con mysql

Finalmente se procede a dar un clic en **OK**, y presenta una ventana figura D.4, donde se puede observar que los datos ya están cargados en la herramienta y están listos para aplicarles cualquier algoritmo de los que posee Weka.

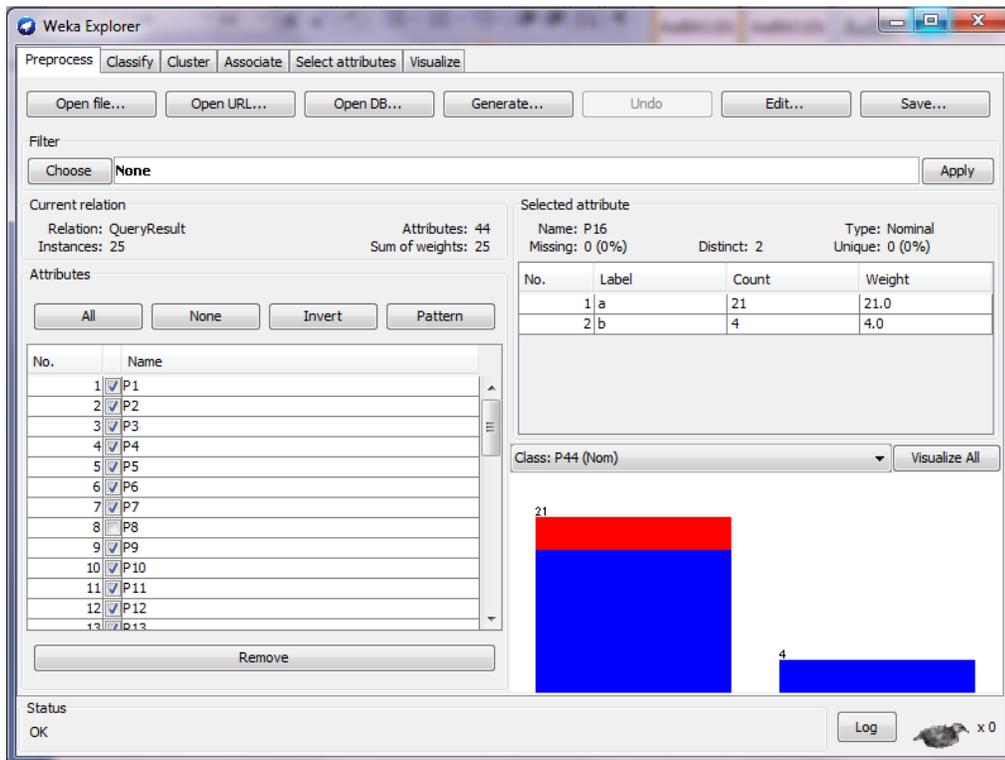


Figura D.4 Datos de los estilos de aprendizaje cargados en Weka.

# ANEXO E

## *Paper*

## APLICACIÓN DE TÉCNICAS DE MINERÍA DE DATOS PARA IDENTIFICAR PATRONES DE COMPORTAMIENTOS RELACIONADOS CON LAS ACCIONES DEL ESTUDIANTE CON EL EVA DE LA UTPL

Priscila Valdiviezo  
Unidad de Virtualización  
[pmvaldiviezo@utpl.edu.ec](mailto:pmvaldiviezo@utpl.edu.ec)

Susana Arias  
IICC IC-IS  
[saarias@utpl.edu.ec](mailto:saarias@utpl.edu.ec)

Marcia Sarango  
Universidad Técnica Particular de Loja  
[mysarangox@utpl.edu.ec](mailto:mysarangox@utpl.edu.ec)

**RESUMEN:** *La presente investigación está enfocada en la identificación de patrones de comportamiento relacionados con las acciones de los estudiantes que utilizan el entorno virtual de aprendizaje (EVA), se realizó un análisis previo de la base de datos correspondiente al periodo académico Abril2011/Agosto2001, para seleccionar exclusivamente las entidades que contienen información útil sobre las acciones que realizan los estudiantes. En la utilización de técnicas de minería se optó por la clasificación: arboles de decisión, reglas de decisión y métodos bayesianos éstas se utilizaron para clasificar los indicadores: participación o interés en el curso y utilización de herramientas por carrera siendo estas informática y abogacía, mientras que el análisis de secuencias, mediante la agrupación o agrupación en clústeres (clustering), fue utilizado para conocer los grupos de los estudiantes con características similares en relación al indicador de la utilización de herramientas y para encontrar el estilo de aprendizaje dominante y conocer cómo aprenden los estudiantes mediante el análisis de los estilos. Con los resultados obtenidos se detalló en un modelo de datos: indicadores, atributos y estilos de aprendizaje del estudiante.*

**PALABRAS CLAVE:** Minería de datos, Moodle, clasificación, clustering.

### 1. INTRODUCCIÓN

En el campo de Tecnologías para la Educación, la adaptación de sistemas educativos ofrece una forma avanzada en un ambiente de aprendizaje que intenta satisfacer las necesidades de los diferentes estudiantes, así como sistemas de construcción de

un modelo de conocimiento del estudiante, las metas y preferencias del mismo. Aquí es donde juega un papel fundamental la aplicación de las técnicas de minería contribuyendo al descubrimiento de patrones de comportamiento de los estudiantes durante su interacción con el EVA de la UTPL. Algunos entornos de aprendizaje virtual (EVA) son adecuados para la investigación del comportamiento de los alumnos en el aprendizaje. Uno de los más populares de éstos es Moodle ampliamente utilizado para la presentación de materiales de aprendizaje, así como para los debates entre los alumnos. Esta herramienta permite a un profesor no sólo informar materiales de aprendizaje de forma flexible, sino que también permite proporcionar la posibilidad de que los alumnos participen en las discusiones comunes, chats sincrónicos, crear sus blogs, archivos de revisión de vídeo de las conferencias, utilizar el correo electrónico, etc. Moodle es también una poderosa herramienta para el seguimiento de las acciones de los estudiantes y la interpretación de éstos resultados. Existe una estrecha relación entre los ámbitos de la Inteligencia Artificial (IA) y Aprendizaje Autónomo contando con las principales fuentes de minería de datos técnicas, procesos y métodos de la educación. En [1], el autor establece las oportunidades de investigación en la IA y la educación en base a tres modelos de los procesos educativos: modelos como herramienta científica, se utilizan como medios para la comprensión y la predicción de algunos aspectos de una situación de la educación; modelos como componente: corresponde a alguna característica de la enseñanza o proceso de aprendizaje y se utiliza como un componente de un artefacto educativo, y modelos base para el diseño de objetos educativos: ayudar al diseño de herramientas informáticas para la

educación, proporcionando metodologías de diseño y los componentes del sistema.

## 2. EL PROCESO DE MINERÍA DE DATOS EN LA ENSEÑANZA A DISTANCIA

Los métodos de aplicación de minería en la educación a distancia consisten en cuatro pasos (Figura 2.1):

1. La acumulación de datos.
2. Preparación de los datos acumulados.
3. Aplicación de métodos de minería de datos para los datos seleccionados.
4. Interpretación y análisis de los datos seleccionados [4]

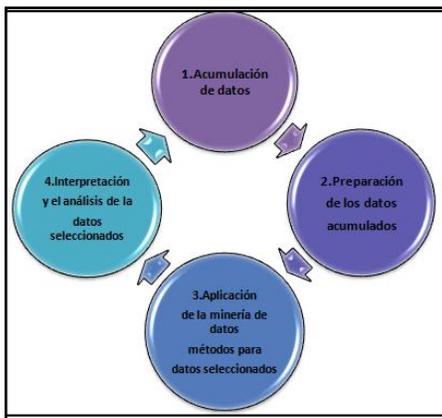


Figura 1: Pasos de los métodos de aplicación de minería de datos en la enseñanza a distancia (Romero, Ventura, y García, 2007)

Pasos que se deben tener en cuenta, para guiarse por una línea de trabajo que permite tener claro que es lo que se desea encontrar con los datos recopilados.

## 3. CARACTERÍSTICAS A TENER EN CUENTA EN LA CONSTRUCCIÓN DE UN MODELO DEL ESTUDIANTE

En [5] se mencionan que las características pueden ser propias al proceso educativo como son el nivel de comprensión de un tema, estilos de aprendizaje, gustos del usuario a la hora de aprender, características psicológicas como el estado de ánimo,

las metas y por último su entorno como el clima y su estrato. Todas estas características relacionadas entre sí forman lo que se llama el modelo del estudiante, lo que permite construir un perfil para acomodar el sistema a las necesidades del usuario.

Para la construcción de un modelo del estudiante se deben tener en cuenta varios aspectos los cuales están especificados en la (figura 2).

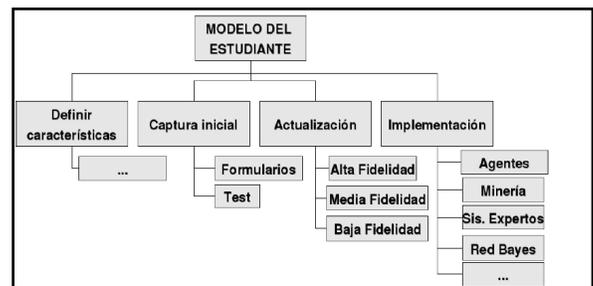


Figura 2: Aspectos a tener en cuenta en la construcción de un modelo del estudiante (Duque y Néstor, 2008)

En consecuencia el proceso de construcción del modelo seguirá la siguiente metodología:

- ❖ Definir las características del estudiante que afecten el proceso de enseñanza.
- ❖ Establecer las técnicas necesarias para realizar la captura inicial de los valores de cada una de las características definidas en el punto anterior.
- ❖ Seleccionar una forma de actualizar dinámicamente los valores de cada una de las características que lo ameriten.
- ❖ Seleccionar las técnicas más adecuadas para implementar el modelo en un sistema de educación virtual adaptativo.

## 4. ACCIONES DE LOS ESTUDIANTES EN EL EVA

A continuación en la siguiente tabla se describe las principales acciones que el estudiante realiza sobre la plataforma.

| module     | action                | DESCRIPCIÓN  |
|------------|-----------------------|--|
| COURSE     | VIEW                  | Número de accesos al curso seleccionado  |
| ASSIGNMENT | VIEW                  | Número de veces que el usuario accede a las tareas enviadas por el profesor.   |
| FORUM      | ADD_POST              | Número de veces que el usuario envía un POST <sup>15</sup> , a un foro es decir un mensaje que da respuesta a alguna pregunta que realizó el profesor. |
| FORUM      | VIEW_DISCUSSION       | Número de veces que el usuario revisa las discusiones dentro de un foro.   |
| FORUM      | VIEW_FORUM            | Número de veces que el usuario revisa o accede a los foros de discusión.   |
| FORUM      | UPDATE_POST           | Número de veces que el usuario actualiza el POST que se envió al foro.   |
| FORUM      | ADD_DISCUSSION        | Número de veces que el usuario agrega un tema de discusión o debate.   |
| RESOURCE   | VIEW                  | Número de veces que el usuario accede descarga recursos 16 subidos por el profesor.  |
| USER       | UPDATE                | Número de veces que el usuario actualiza los datos de su perfil.   |
| USER       | VIEW                  | Número de veces que el usuario revisa o accede a su perfil.  |
| UPLOAD     | UPLOAD                | Número de veces que el usuario sube una tarea.   |
| MESSAGE    | WRITE                 | Número de mensajes que el usuario ha escrito.  |
| MESSAGE    | HISTORY               | Número de visitas al historial de los mensajes.  |
| QUIZ       | VIEW                  | Número de veces que el usuario revisa o accede a los cuestionarios.  |
| QUIZ       | TIMESTART, TIMEFINISH | Tiempo que tarda en la resolución de un cuestionario.  |
| QUIZ       | ATTEMPT               | Número de veces que el usuario intenta resolver un cuestionario.   |

Tabla 1: Atributos de las acciones del usuario sobre la plataforma

## 5. DESCRIPCIÓN DEL ESCENARIO

La base de datos para la búsqueda de patrones, contiene información de todos los cursos, en las modalidades (Clásica y Abierta) de estudio en la UTPL del periodo correspondiente a

<sup>15</sup> POST se refiere a la acción de enviar un mensaje hacia el servidor principal.

Abril2011/Agosto2011, conviene indicar que lo que interesa es la información de los estudiantes de la modalidad Abierta la cual será el foco del desarrollo de la tesis para una representación de indicadores en un modelo de estudiante en base a los patrones de comportamiento en relación con las acciones. Para este estudio se tomará como muestra los cursos detallados en la tabla 2, con el objeto de tener una muestra representativa que permita identificar diferentes comportamientos de estudiantes de algunas carreras, en base a estos comportamientos, se determinaran atributos e indicadores que pueden ser representados en un modelo de estudiante para la UTPL, que considere información que describa en gran medida las acciones que los estudiantes de la UTPL realizan sobre la plataforma, es por ello que en este caso se ha creído conveniente abarcar al menos dos carreras.

| CARRERA           | ÁREA              | CURSO  | CANTIDAD DE ESTUDIANTES |
|-------------------|-------------------|--|-------------------------|
| Informática       | Técnica           | LOGICA DE LA PROGRAMACION                                  | 205                     |
|                   |                   | MATEMATICAS DISCRETAS                                      | 142                     |
|                   |                   | INTELIGENCIA ARTIFICIAL                                    | 18                      |
|                   |                   | SEMINARIO II (Desarrollo de páginas web dinámicas con php) | 23                      |
|                   |                   | <b>Total: 388</b>  |                         |
| Abogacía          | Socio Humanística | MEDICINA LEGAL   | 119                     |
|                   |                   | ETICA Y DERECHOS HUMANOS                                   | 156                     |
|                   |                   | LEGISLACION MERCANTIL MONETARIA Y BANCARIA                 | 122                     |
| <b>Total: 397</b> |                   |  |                         |

Tabla 2: Muestra de las carreras con sus respectivos cursos

### 5.1 ALGORITMOS PARA DETERMINAR LOS COMPORTAMIENTOS DEL ESTUDIANTE EN EL EVA.

Una de las técnicas que se empleara en la minería para determinar el comportamiento de los estudiantes en base a las acciones que éste realiza sobre el EVA de la UTPL, es la de clasificación, como son los árboles de decisión, reglas de decisión y métodos bayesianos, éstos son utilizados para el indicador de la participación del estudiante en el curso ya que según lo analizado ayudan a predecir una o más variables discretas, basándose en otros atributos del conjunto de datos, el algoritmo hace predicciones, mientras que en el análisis de secuencias, el algoritmo de clústeres de secuencia permite explorar los datos que contienen eventos que pueden vincularse mediante rutas o secuencias. El algoritmo encuentra las secuencias más comunes mediante la agrupación, o agrupación en clústeres, de las secuencias que son idénticas. Este último método de clústeres es utilizado para el indicador de la utilización de las herramientas analizando de manera individual cada atributo que representan las acciones de los estudiantes para de esta manera proceder a establecer los grupos y determinar cual hace más uso de determinada herramienta dentro de la plataforma.

| Algoritmos  | Opción      |
|---|-------------|
| <ul style="list-style-type: none"> <li>✓ C4.5 o J48</li> <li>✓ REPtree</li> </ul>         | Trees       |
| <ul style="list-style-type: none"> <li>✓ BayesNet</li> <li>✓ NaiveBayes</li> </ul>        | Classifiers |
| <ul style="list-style-type: none"> <li>✓ JRip</li> </ul>                                  | Rules       |
| <ul style="list-style-type: none"> <li>✓ SimpleKMeans</li> <li>✓ FarthestFirst</li> </ul> | Clústeres   |

Tabla 3: Algoritmos para experimentación

## 5.2 ANÁLISIS DE LOS DATOS

En esta dimensión se enfocará concretamente a la interacción personal del estudiante con la plataforma y es usada para conocer los comportamientos de los estudiantes, según las acciones que realizan en el eva, previo a esto se realiza un análisis de los atributos detectados en cada curso detallados en tabla 1 que en realidad sirven, ya que a partir de éstos se derivan algunas variables que pueden ser incluidas las cuales son:

- ❖ número de accesos al curso.

- ❖ número de accesos a las tareas enviadas por el profesor
- ❖ número de veces que revisa un cuestionario
- ❖ número de veces que descarga un recurso.
- ❖ número de veces que el usuario revisa su perfil.
- ❖ número de veces que el usuario accede a un foro.
- ❖ número de veces que contesta un foro.
- ❖ número de veces que revisa una discusión dentro de un foro.
- ❖ número de veces que actualiza un post dentro del foro.
- ❖ número de veces que agrega un tema de discusión o debate.
- ❖ número de veces que actualiza el perfil de usuario.
- ❖ número de mensajes que envía el usuario
- ❖ número de visitas al historial de los mensajes.
- ❖ número de veces que envía o sube una tarea.
- ❖ número de mensajes (twitter).
- ❖ número de intentos en la evaluación a distancia del primer bimestre.
- ❖ tiempo que tarda en la evaluación a distancia del primer bimestre.
- ❖ número de intentos en la evaluación a distancia del segundo bimestre.
- ❖ tiempo que tarda en la evaluación a distancia del segundo bimestre.
- ❖ Tiempo promedio de acceso al curso

La interacción del usuario con el sistema es seguida a través de los logs. Los recursos (materiales) son ayudas digitales que contienen información que puede ser leída, vista o bajada para extraer información que apoye el proceso de aprendizaje, éstos permiten conocer el material que ha visto el estudiante.

## 5.3 INDICADORES CORRESPONDIENTES A LAS ACCIONES DEL ESTUDIANTE CON EL EVA.

Un **indicador**, según algunas referencias es descrito como: "Herramientas para clarificar y definir de, forma más precisa, objetivos e impactos y son

medidas verificables de cambio o resultado diseñadas para contar con un estándar contra el cual evaluar, estimar o demostrar el progreso con respecto a metas establecidas, facilitan el reparto de insumos, produciendo productos y alcanzando objetivos”<sup>17</sup>. Por lo que conociendo de antemano todas las características que sirvieron para encaminar según con los datos que contamos y partiendo desde los objetivos, es decir, que es lo que buscamos lograr al definir los indicadores que corresponden a las acciones que los estudiantes realizan, se ha creído conveniente establecer los siguientes indicadores:

- ✓ Participación o interés en el curso.
- ✓ Utilización de herramientas.

**Participación o interés en el curso:** Este atributo representa el interés que el estudiante tiene en determinado curso midiendo la aportación que este haya realizado en la plataforma, datos útiles para poder sugerir sobre contenidos a los estudiantes que presenten un nivel de interés bajo, y contenidos un poco más complejos para los estudiantes que presenten un nivel de participación alto. En [2], presentan un apartado en el que describen el tipo de interacción del usuario, esto ha servido como base para delimitar los niveles en cada indicador que se ha establecido para efectos de esta tesis, ya que en el trabajo que ellos realizaron muestran la interacción conformista tipificándola como bajo, interacción consiente como medio e interacción autónoma como alto, lo que da pie y se cree conveniente que para este trabajo, en cada uno de los indicadores que presenten valores de incidencia alto será denotado como permanente (P), valores de incidencia medio como moderado (M) y los valores de incidencia bajo como escaso (E). Quedando así establecidos los siguientes niveles:

- ✓ Permanente (P)
- ✓ Moderado (M)
- ✓ Escaso (E)

**Utilización de herramientas:** Aquí se trata de las acciones que el estudiante realiza sobre las herramientas tales como foros, mensajes, recursos, tareas, twitter, y cuestionario, puesto que son actividades subidas por el profesor en la plataforma para el aprendizaje del alumno. Para este indicador primero se detalla las acciones por ejemplo para el caso de los foros tenemos algunas acciones como son: ver foro, ver discusión dentro de un foro, añadir discusión, añadir post y actualizar post, estas acciones sobre la respectiva herramienta se encuentran detallado en la Tabla 4. Aquí también utilizaremos los siguientes niveles:

- ✓ Permanente (P)
- ✓ Moderado (M)
- ✓ Escaso (E)

Para representar el nivel de utilización de herramientas que tiene el estudiante sobre la plataforma.

| Herramienta  | Acción                                       |
|--------------|--|
| foros        | Ver foro                                     |
|              | Añadir post                                  |
|              | Ver discusión dentro del foro                |
|              | Actualizar post                              |
|              | Añadir tema de discusión o debate            |
| recursos     | Descargar archivos                           |
| tareas       | Visualizar tareas propuestas por el profesor |
|              | Enviar tareas                                |
| twitter      | Enviar mensaje twitter                       |
| mensajes     | Escribir mensajes                            |
|              | Ver historial de mensajes                    |
| cuestionario | Responder un cuestionario                    |
|              | Visualizar preguntas                         |

Tabla 4: Acciones en las herramientas utilizadas en los cursos

La evaluación de éstos niveles de seguimiento tiene su fundamento en la utilización de indicadores ya que los datos obtenidos permiten la acción tutorial cuando se detecta una desatención del alumno de sus actividades académicas, ya que según la participación del alumno en el entorno virtual, y con los atributos mencionados anteriormente se podrá dar soporte acertado en cuanto a contenidos en los cursos que se mencionan en la descripción del escenario sección 5.

## 5.4 ESCENARIO DE EXPERIMENTACIÓN

Los algoritmos y datos que se utilizaron en las experimentaciones que se realizaron con la

<sup>17</sup> Organización de las Naciones Unidas (ONU). Integrated and coordinated implementation and follow-up of major. United Nations conferences and summits. Nueva York, Estados Unidos de América, 10 y 11 de mayo de 1999, p. 18. Recuperado el 1 de Junio 2011, de <http://bit.ly/qkyoxM>.

herramienta WEKA, dentro de la minería de datos se detallan a continuación:

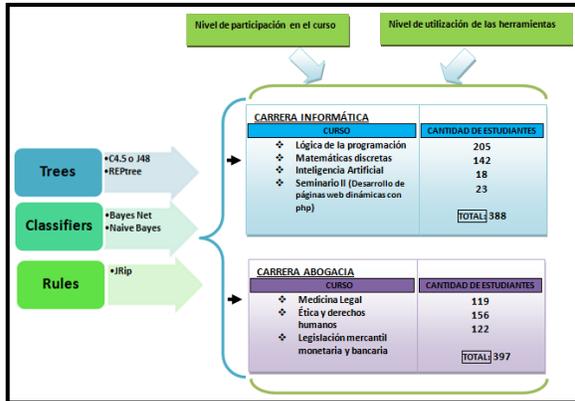


Figura 3: Escenario para el nivel de participación en el curso y el nivel de utilización de las herramientas por carrera.

En la figura 3, se puede apreciar los algoritmos C4.5 o J48 (Trees), Bayes Net y Naive Bayes (Classifiers o métodos bayesianos) y JRip (Rules o reglas de decisión) para la experimentación, los cuales forman parte de la técnica de clasificación, estos serán aplicados a cada una de las carreras para encontrar dos indicadores: el nivel de participación en el curso y el nivel de utilización de las herramientas, en el caso de informática se cuenta con 388 instancias o cantidad de estudiantes mientras que en abogacía 397 instancias, a partir de la experimentación se seleccionará el algoritmo que presente los mejores resultados para ser sujetos a análisis e interpretación y posteriormente proceder a realizar comparaciones entre los resultados obtenidos de ambas carreras ver tabla 8, previo a esto se realizó una descripción de los algoritmos que se utilizarán (en el punto 5.1), y un análisis de los datos de cada curso de ahí se obtuvieron algunas variables ver (en el punto 5.2).

Las variables que se presentan son los datos a extraer de cada curso, mediante la utilización de sentencias SQL ver en anexo B.

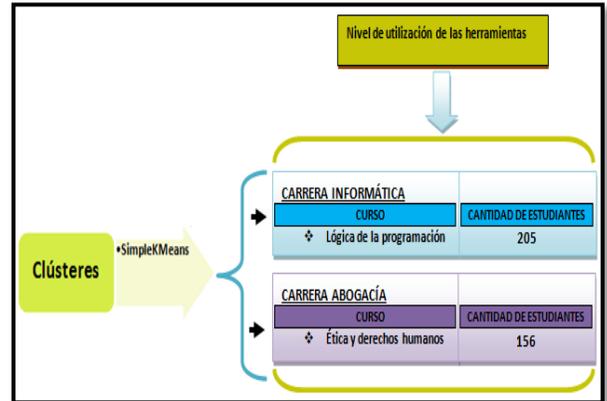


Figura 4: Escenario para el nivel de utilización de las herramientas por curso

En la figura 4, se presenta un esquema de la aplicación del algoritmo SimpleKMeans (Clústeres) para la experimentación, este algoritmo será aplicado a un curso de cada carrera, para encontrar el indicador del nivel de utilización de las herramientas por curso, ya no por carrera como el escenario de la figura 3, esta vez se selecciona el que tenga la mayor cantidad de estudiantes o instancias, ya que al contar con un número de instancias considerable, los resultados obtenidos son más confiables, los cursos resultantes son: Lógica de la programación de la carrera de Informática con 205 instancias y, ética y derechos humanos de Abogacía con 156 instancias.

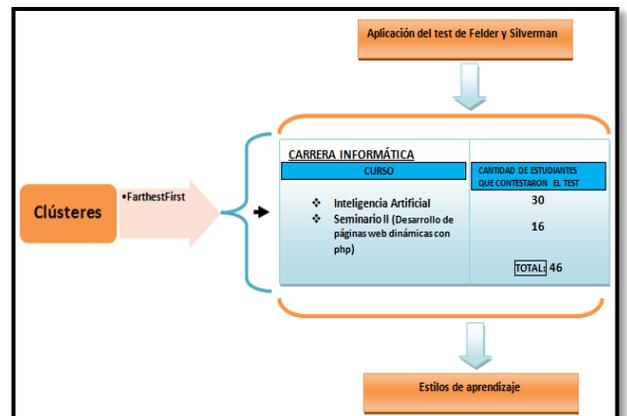


Figura 5: Escenario para encontrar el estilo de aprendizaje dominante

En la figura 5, se observa el algoritmo FarthestFirst (Clústeres), el cual se aplicó a los cursos de los últimos ciclos de la carrera de

informática: Inteligencia Artificial y Seminario II (Desarrollo de páginas web dinámicas con php) para encontrar el estilo de aprendizaje predominante y realizar un análisis de cómo aprenden los estudiantes en esta carrera, previa a la aplicación del algoritmo se habilita a manera de enlace, el test de Felder y Silverman a los estudiantes en cada curso quedando como resultado un total de 46 estudiantes que dieron contestación al test.

## 6. LIMPIEZA Y TRANSFORMACIÓN DE DATOS

La recopilación de datos debe ir acompañada de una limpieza e integración de los mismos, para que estos estén en condiciones para su análisis. Los beneficios del análisis y de la extracción de conocimiento a partir de datos dependen, en gran medida, de la calidad de los datos recopilados. Además, generalmente, debido a las características propias de las técnicas de minería de datos, es necesario realizar una transformación de los datos para obtener una “materia prima” que sea adecuada para el propósito concreto y las técnicas que se quieren emplear. En definitiva, el éxito de un proceso de minería de datos depende, no solo de tener todos los datos necesarios (una buena recopilación), sino de que estén íntegros, completos y consistentes (una buena limpieza e integración) [3].

En esta fase se ha estimado conveniente la transformación del tiempo de la tabla `mdl_quiz_attempts` descrita en el Capítulo 2, ya que originalmente en la base de datos se encontraba en un formato no legible para humanos esta es `timestamp`, haciendo uso de la función `select from_unixtime();` esto nos facilita las unidades para saber el tiempo que un estudiante tarda en la resolución de un cuestionario. Consolidando éstos datos apropiadamente para continuar con la siguiente fase. Para la parte de integración de las distintas fuentes de datos registradas primero en Excel, se procedió a crear tablas en una base de datos de las asignaturas aquí recopilaremos todos los atributos que nos serán de utilidad.

Con la información obtenida a través consultas SQL que para recopilar todas las acciones que se

realizaron en determinado curso, entre las actividades de pre procesamiento se aplicó limpieza de los datos en el momento de ejecutar la consulta ya que en el caso de que hubieran existido estudiantes que constaban como participantes del curso pero sin embargo nunca ni siquiera ingresaron al curso en el EVA, por ende éstos no realizaron ninguna acción y no fueron tomados en cuenta, también se eliminaron algunas inconsistencias a través de correcciones manuales, tales como la existencia de espaciados innecesarios. Adicional a esto se realizó otra consulta para las acciones en los cuestionarios, tales como el número de intentos y tiempo que emplea el estudiante para rendir el módulo antes mencionado. Para efecto del registro de cada una de las acciones, se lo alojó en una hoja de cálculo Excel, posteriormente se alojó a la base de datos minería acciones para luego ser cargado en WEKA.

## 7. DISCRETIZACIÓN

La discretización, es la conversión de un valor numérico en un valor nominal ordenado el cual representa un intervalo o “bin”, cuyo proceso consiste en la búsqueda de los intervalos más adecuados para cada dominio numérico, para esto lo primero que se realizó fue conectar desde WEKA<sup>18</sup> nuestra base de datos donde recopilamos información necesaria para la extracción del conocimiento, misma base que antes fue alojada en el manejador de base de datos `SQLyog8.05` mediante sentencias SQL mismas que se guardaron en un `.txt` o archivo de texto plano. Seguido esto con la ayuda de una consulta (Figura 6) se extrajo a la aplicación los atributos que servirán para determinar la participación en el curso mismos atributos que se procederá a discretizar desde WEKA.

---

<sup>18</sup> WEKA (Waikato Environment for Knowledge Analysis - Entorno para Análisis del Conocimiento de la Universidad de Waikato) <http://www.cs.waikato.ac.nz/ml/weka/>

```

1 SELECT DISTINCT
2 informatica.userid,
3 informatica.num_accesos_curso,
4 informatica.tavg_acceso_curso/num_accesos_curso AS tavg_acceso_curso,
5 informatica.num_accesos_tareas,
6 informatica.num_veces_sube_tarea,
7 informatica.num_veces_descarga_recurso,
8 informatica.num_veces_revisa_cuestionario,
9 informatica.num_intevaladpb,
10 informatica.tie_evaladpb/num_intevaladpb AS tie_evaladpb,
11 informatica.num_intevaladsb,
12 informatica.tie_evaladsb/num_intevaladsb AS tie_evaladsb,
13 informatica.num_msn_escritos,
14 informatica.num_veces_revisa_historial_msn,
15 informatica.num_veces_revisa_perfil,
16 informatica.num_accesos_foro,
17 informatica.num_veces_contesta_foro,
18 informatica.`num_veces_revisa_discusión_foro`,
19 informatica.num_veces_actualiza_contest_foro,
20 informatica.`num_veces_agrega_tema_discusión`,
21 informatica.num_veces_actualiza_perfil,
22 informatica.num_msn_twitter,
23 informatica.`participación_curso`
24 FROM
25 informatica
26 GROUP BY
27 informatica.userid,
28 informatica.num_accesos_curso,
29 informatica.num_accesos_tareas,
30 informatica.num_veces_sube_tarea,
31 informatica.num_veces_descarga_recurso,
32 informatica.num_veces_revisa_cuestionario,
33 informatica.num_intevaladpb,
34 informatica.num_intevaladsb,
35 informatica.num_accesos_foro,

```

Figura 6: Consulta SQL para extraer atributos de la participación en el curso (informática).

Ahora bien una vez que ya tenemos determinados cuales van a ser nuestros intervalos se procede a través de WEKA a evaluar la participación del estudiante en el curso, según [6] denota como una técnica de minería de datos en la clasificación, a las técnicas basadas en conteo de frecuencias, que se apoyan en contar la frecuencia con la que dos o más sucesos se presentan, lo que ha permitido incluir y dejar establecido que en este trabajo de tesis se ha creído conveniente tomar en cuenta la moda<sup>19</sup> en cada atributo y por cada fila de nuestro registro de datos como se muestra en la (Figura 6) iremos asignando el nivel ya sea Permanente (P), Moderado (M) y Escaso (E). Éstos pasos se realizan por cada curso los que se detallan aquí son los referentes a la carrera de Informática.

<sup>19</sup>Moda, es el valor que cuenta con una mayor frecuencia en una distribución de datos. Se puede hallar la moda para variables cualitativas y cuantitativas. <http://bit.ly/pkxuQ8>

| Relación: QueryResult               | 19: num_veces_agrega_tema_discusión | 20: num_veces_actualiza_perfil | 21: num_msn_twitter | 22: participación_curso |
|-------------------------------------|-------------------------------------|--------------------------------|---------------------|-------------------------|
| 19: num_veces_agrega_tema_discusión | Numeric                             | Numeric                        | Numeric             | Nonnal                  |
| 6.0                                 | 2.0                                 | 4.0                            | 3.0                 | M                       |
| 0.0                                 | 0.0                                 | 0.0                            | 0.0                 | E                       |
| 0.0                                 | 0.0                                 | 0.0                            | 0.0                 | M                       |
| 0.0                                 | 0.0                                 | 0.0                            | 8.0                 | P                       |
| 0.0                                 | 2.0                                 | 1.0                            | 2.0                 | M                       |
| 0.0                                 | 2.0                                 | 0.0                            | 0.0                 | E                       |

Figura 7: Asignación de la clase para el indicador del nivel de participación.

Este proceso se lo realiza de igual manera para el nivel de utilización de herramientas en las dos carreras, aquí se indica en general tomando como ejemplo a la carrera de informática que en efecto fue sujeta a este paso importante la *discretización*.

## 8. MINERIA DE DATOS

En la minería de datos, es necesario tener los datos, ya generados como atributos, además de que en esta fase se evaluarán técnicas y algoritmos de aprendizaje automático para elegir cuál será el más apropiado para los datos que se ha seleccionado. El proceso de la minería será efectuado a través de la herramienta, de minería de datos WEKA.

Dentro de este proyecto de tesis se mencionó que los atributos a identificar del modelo de estudiante atenderán a un modelo predictivo de las instancias, y para este tipo de modelo se optara por utilizar las tareas que son clasificación y clustering.

Una vez que se ha configurado en WEKA correctamente la moda del nivel de participación en el curso, guiándonos en los intervalos que con la ayuda de la misma herramienta se obtuvieron y para evitar posibles confusiones primero se tabuló dicha asignación en una hoja de cálculo para luego proceder a situar el nivel de participación en el curso según el número mayor de frecuencia o el mayor número de veces que se repite la clase E, M y P, ese será el valor cualitativo a asignar, ver en (Figura 7).

Terminado esto se procedió a hacer uso de los algoritmos que posee WEKA y se guardó los modelos

generados por cada experiencia de algoritmo utilizada ya, que luego las matrices de confusión también serían sujetas a análisis. La técnica de evaluación utilizada fue la de cross-validation (validación cruzada) esta consiste en que: dado un número n se divide los datos en n partes y, por cada parte, se construye el clasificador con las n-1 partes restantes y se prueba con esa parte. Así tenemos que por cada una de las n particiones para construir un modelo con todos los datos se obtienen sus ratios de error y precisión, promediando los n ratios de error [3].

## 9. RESULTADOS

A continuación se muestran los resultados de los algoritmos que presentaron los mejores resultados, siendo escogidos en base a la mejor precisión en la clasificación de instancias.

### INFORMÁTICA

| Indicador                            | Algoritmo Seleccionado |
|--------------------------------------|------------------------|
| Nivel de participación en el curso   | REPTree                |
| Nivel de utilización de herramientas | REPTree                |

Tabla 5: Algoritmo seleccionado para cada indicador

### ABOGACÍA

| Indicador                            | Algoritmo Seleccionado |
|--------------------------------------|------------------------|
| Nivel de participación en el curso   | J48                    |
| Nivel de utilización de herramientas | J48                    |

Tabla 6: Algoritmo seleccionado para cada indicador

| Nivel de participación en el curso                                     |  | Nivel de utilización de las herramientas                               |  |
|--|--|--|--|
| Informática  | Abogacía   | Informática  | Abogacía   |
| a b c <-- Classified as<br>0 0 1   a=P<br>0 0 8   b=M<br>0 0 379   c=E | a b c <-- Classified as<br>0 0 0   a=P<br>0 0 2   b=M<br>0 0 995   c=E | a b c <-- Classified as<br>0 0 0   a=P<br>0 0 9   b=M<br>0 1 378   c=E | a b c <-- Classified as<br>0 0 0   a=P<br>0 0 2   b=M<br>0 0 995   c=E |

Tabla 7: Comparación de las matrices de confusión entre informática y abogacía

❖ **Nivel de participación en el curso:** En la carrera de informática se nota que de los 379 estudiantes que presentan un nivel de participación en el curso Escaso (E), todos son correctamente seleccionados, en el caso del nivel Moderado (M) presenta la cantidad de 8 estudiantes que son, los que tienen un nivel medio de participación o interés en el curso, pues resulta ser que en informática existen más estudiantes que participan en el curso en un nivel Moderado, que los de la carrera de abogacía ya que aquí son apenas 2 los que presentan esta incidencia y para el nivel Permanente (P) tenemos que, en informática se tiene la representación de 1 solo estudiante, mientras que para la carrera de abogacía no existe representación alguna observando que la cantidad de estudiantes que acogen este nivel es igual a 0, en base a todas estas observaciones, se puede concluir en general que el nivel predominante en los estudiantes es el Escaso, tanto para los de la carrera de informática como para los de abogacía, ya que en ambos casos la gran mayoría de estudiantes se encuentran concentrados en ese nivel.

❖ **Nivel de utilización de las herramientas:** Aquí se nota que en la carrera de informática de los 379 estudiantes que presentan un nivel de participación en el curso Escaso (E), 378 son clasificados correctamente y 1 con error, en lo que concierne al nivel Moderado (M) se aprecia que existen 9 estudiantes que hacen uso de las herramientas en un nivel medio, caso contrario al de los estudiantes de la carrera de abogacía ya que solamente 2 estudiantes presentan actitudes en éste nivel mientras que para el nivel

Permanente (P), ambas carreras se encuentran en iguales condiciones ya que tienen una incidencia de estudiantes igual a 0. En base a lo expuesto, se puede concluir de manera general que los estudiantes de la carrera de informática como de abogacía presentan mayor incidencia en el nivel de utilización de las herramientas Escaso.

### NIVEL DE UTILIZACIÓN DE HERRAMIENTAS

#### RESULTADOS CON SimpleKMeans

Con el algoritmo caracterizado por su sencillez SimpleKMeans además de ser el más utilizado para hacer clustering y al extenso material bibliográfico con el que se cuenta sobre este algoritmo, teniendo en cuenta que es necesario realizar varios experimentos con este para obtener el resultado más óptimo posible, hacemos uso de éste ya que nos interesa evaluar cada atributo que conforma el nivel de utilización de herramientas para lo cual hemos elegido dos grupos de población, que hacen referencia a dos asignaturas en específico Lógica de la programación y Ética y derechos humanos a estas asignaturas se les aplicará el algoritmo, utilizando los valores reales obtenidos para cada atributo que conforman este indicador del nivel de utilización de las herramientas de la plataforma conforme las acciones que éstos realizan sobre las mismas, de tal manera para posteriormente poder proceder a realizar las comparaciones pertinentes entre estas dos asignaturas de la carrera de Informática.

| Nivel de utilización de las herramientas |                  |                           |            |                          |            |
|--|------------------|---------------------------|------------|--------------------------|------------|
| Herramienta                              | Grupos - Clúster | Lógica de la programación |            | Ética y derechos humanos |            |
|  |                  | Instancias                | Nivel      | Instancias               | Nivel      |
| Foros                                    | 0                | 18%                       | Permanente | 26%                      | Permanente |
|  | 1                | 40%                       | Moderado   | 14%                      | Escaso     |
|  | 2                | 42%                       | Escaso     | 60%                      | Moderado   |
| Recursos                                 | 0                | 18%                       | Permanente | 26%                      | Moderado   |
|  | 1                | 40%                       | Moderado   | 14%                      | Permanente |
|  | 2                | 42%                       | Escaso     | 60%                      | Escaso     |
| Tareas                                   | 0                | 18%                       | Permanente | 26%                      | Permanente |
|  | 1                | 40%                       | Moderado   | 14%                      | Moderado   |
|  | 2                | 42%                       | Escaso     | 60%                      | Escaso     |
| Mensajería                               | 0                | 18%                       | Permanente | 26%                      | Permanente |
|  | 1                | 40%                       | Moderado   | 14%                      | Moderado   |
|  | 2                | 42%                       | Escaso     | 60%                      | Escaso     |
| Twitter                                  | 0                | 18%                       | Permanente | 26%                      | Moderado   |
|  | 1                | 40%                       | Moderado   | 14%                      | Permanente |
|  | 2                | 42%                       | Escaso     | 60%                      | Escaso     |
| Cuestionario                             | 0                | 18%                       | Permanente | 26%                      | Permanente |
|  | 1                | 40%                       | Moderado   | 14%                      | Moderado   |
|  | 2                | 42%                       | Escaso     | 60%                      | Escaso     |

Tabla 8: Comparación de grupos

Los estudiantes que presentan más interacción con la herramienta (foros) pertenecen al curso de ética y derechos humanos, mientras que en los estudiantes del curso de lógica de la programación presentan más interacción con las herramientas (foros, recursos, tareas, mensajería, twitter y cuestionario). Denotando claramente a nivel general que de entre los dos cursos, la utilización de las herramientas permanente y moderado, representan mayor interacción los estudiantes que corresponden al curso lógica de la programación (informática).

**ESTILOS DE APRENDIZAJE**

**RESULTADOS CON FarthestFirst**

La Población de estudio fue conformada por 46 estudiantes que fueron los que dieron contestación al test, y que corresponden a los últimos ciclos de la carrera de "Informática" comprendiendo las asignaturas de: Seminario II(Desarrollo de páginas web dinámicas con php) e Inteligencia artificial, de la UTPL modalidad Abierta, se les realizó el test de Felder ver en anexo C, este test fue habilitado en el EVA a manera de enlace y posteriormente los resultados obtenidos se fueron almacenado en una base de datos, para proceder a realizar luego lo que es la extracción de conocimiento.

| Grupo 1 - Clúster 0   |              |
|-----------------------|--------------|
| Estilo de Aprendizaje | Concurrencia |
| I (Intuitivo)         | 4            |
| V (Visual)            | 7            |
| AC (Activo)           | 4            |
| SC (Secuencial)       | 4            |
| SN (Sensorial)        | 8            |
| AU (Auditivo)         | 3            |
| R (Reflexivo)         | 7            |
| G (Global)            | 5            |
| Total = 42            |              |

Tabla 9: Resultados de correspondencia-clúster 0

Se puede concluir que es un grupo conformado por el 48%, de la población encuestada que manifiesta mayor tendencia para aprender en el estilo **visual-sensorial (V-SN)**, **reflexivo-global (R-G)**, denotando que son estudiantes que aprenden mejor trabajando individualmente, observando siendo cuidadosos y creativos en la forma que realizan sus tareas familiarizándose mejor con imágenes y gráficos para entender todo sobre un tema.

| Grupo 2 - Clúster 1   |              |
|-----------------------|--------------|
| Estilo de Aprendizaje | Concurrencia |
| I (Intuitivo)         | 4            |
| V (Visual)            | 4            |
| AC (Activo)           | 8            |
| SC (Secuencial)       | 6            |
| SN (Sensorial)        | 8            |
| AU (Auditivo)         | 6            |
| R (Reflexivo)         | 3            |
| G (Global)            | 3            |
| Total = 42            |              |

Tabla 10: Resultados de correspondencia-clúster 1

Se puede concluir que se trata de un grupo conformado por el 52%, de la población encuestada, que presenta mayor tendencia para aprender en el estilo **activo-secuencial (AC-SC)**, **sensorial-auditivo (SN-AU)**, reflejando que son estudiantes que tienen mayor facilidad para trabajar en equipo expresando sus inquietudes, discutiéndolas y escuchando las opiniones de los demás, para luego sacar soluciones concretas que le sean de utilidad.

**10. REPRESENTACIÓN DE INDICADORES Y ATRIBUTOS DEL ESTUDIANTE EN UN MODELO DE DATOS.**

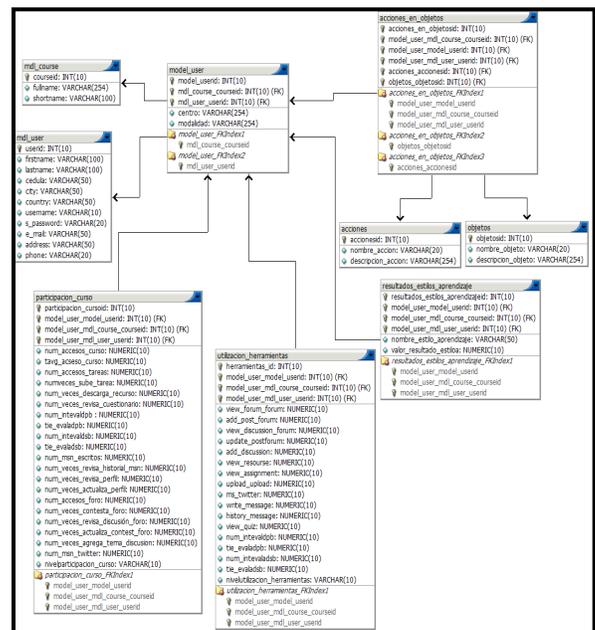


Figura 8: Modelo de datos del estudiante

El modelo de datos del estudiante se presenta en la figura 8, donde se realiza ya la unificación de las entidades, atributos y relaciones principales que intervienen, las cuales se han ido describiendo tras el proceso de aplicar la metodología entidad – relación.

## 11. CONCLUSIONES

A partir de las experimentaciones con los datos recogidos y los resultados obtenidos se puede concluir lo siguiente:

- ❖ Con la minería de datos, a partir de la información se descubre y analiza nuevo conocimiento útil, para que un sistema cuente con una representación del estado actual del usuario, que en este caso es la descripción de las acciones del estudiante, representado en un modelo que se obtiene a partir de la personalización y análisis de datos, que son aplicaciones propias del modelado de usuario.
  - ❖ Para predecir el nivel de participación en el curso y el nivel de utilización de las herramientas, en informática y abogacía después de la aplicación de algoritmos de clasificación, el que presentó los mejores resultados fue el REPTree en informática y J48 en abogacía, y con fundamento en las matrices de confusión presentadas por éstos algoritmos se pudo apreciar notablemente que en ambas carreras existe mayor cantidad de estudiantes que presentan una escasa, participación en el curso, así como en la utilización de herramientas.
  - ❖ Se realizó algunas experimentaciones con clustering a los datos de los cursos: lógica de la programación (informática), y ética y derechos humanos (abogacía), el algoritmo seleccionado fue SimpleKMeans para determinar grupos de estudiantes con comportamientos similares. En relación al indicador del nivel de utilización de las herramientas, se encontró que para lógica de la programación el uso de las herramientas foros, recursos, tareas,
- mensajería, twitter y cuestionario, es en un nivel permanente (alto) y moderado (medio), mientras que en ética y derechos humanos los estudiantes presentan un nivel bajo de utilización de las herramientas a excepción de los foros dónde se pudo observar que existe un nivel permanente (alto) de interacción y en algunos casos moderado.
- ❖ Haciendo una comparación comparación de los grupos generados en ambas materias, en los resultados se encontró que los estudiantes del curso de lógica de la programación presentan mayor interacción en la mayoría de las herramientas siendo estas: Foros, recursos, tareas, mensajería, twitter y cuestionario, mientras que los estudiantes de ética y derechos humanos revelan mayor interacción en una sola herramienta la cual es foros, por lo que se deduce que a lo mejor en este curso el profesor habilitó para el estudiante más actividades a desarrollar en esta herramienta, que en las demás, demostrando que la cantidad de utilización de las herramientas depende de las actividades que habilite el profesor en ellas.
  - ❖ En la herramienta cuestionario, se detectó un grupo de estudiantes que cuando utilizan una mayor cantidad de tiempo por resolver la evaluación a distancia, el número de intentos que realizan es menor, que puede ser porque leen detenidamente las preguntas para proceder a dar contestación, en lugar de hacer varios intentos en menor tiempo que son causa y que pueden ser origen de inseguridad en la respuestas a ingresar.
  - ❖ En la búsqueda de los estilos de aprendizaje también se realizaron experimentaciones con clustering, se aplicó el algoritmo FarthestFirst, encontrando dos grupos: en el primer grupo se ubican los estudiantes que presentan mayor tendencia en el estilo visual-sensorial (V-SN), reflexivo-global (R-

- ❖ G), reflejando mayor complacencia para trabajar en sus tareas con la ayuda de gráficos, siendo muy observadores y creativos, aprendiendo mejor trabajando solos, en el segundo grupo están los estudiantes que tienen mayor incidencia en el estilo activo-secuencial (AC-SC), sensorial-auditivo (SN-AU), denotando que receptan mejor la información a través de debates con otras personas, exponiendo diferentes puntos de vista y tienen mayor gusto y predisposición para trabajar en grupo.

## 12. TRABAJOS FUTUROS

Considerar, para trabajos futuros la incorporación de nuevos indicadores que representen acciones de los estudiantes en otras herramientas del EVA que no hayan sido consideradas en este trabajo, que a lo mejor hayan sido identificados en otros proyectos afines:

- ❖ La agregación de las acciones que el estudiante realiza en la herramienta chat, con el objeto de descubrir indicadores de colaboración, debido a que es un recurso de comunicación entre los participantes de un curso.
- ❖ Establecer indicadores de navegabilidad mediante el seguimiento de la ruta que el estudiante efectúa, desde que accede al entorno virtual de aprendizaje, hasta su salida del EVA.
- ❖ Desarrollar un análisis, si el nivel de interacción del estudiante con las herramientas influye en su rendimiento académico con la finalidad de obtener indicadores de rendimiento académico.

## 13. REFERENCIAS

- [1] Baker, M. (2000). *The roles of models in Artificial Intelligence and Education*, *International Journal of Artificial Intelligence in Education*, GRIC-COAST, CNRS & Université Lumière Lyon. Mendés, France.
- [2] Alberca, G., Loja, F., Valdiviezo, P. Jiménez, C., y Agila, M. (2008). Recolección de datos de interacción de alumnos en una plataforma, Universidad Técnica Particular de Loja. Loja, Ecuador.
- [3] Hernández, J., Ramírez, M., y Ferri, C. (2008). *Introducción a la minería de datos*. España: Pearson Education.
- [4] Romero, C., Ventura, S., y García, E. (2007). *Data mining in course management systems: Moodle case study and tutorial*. University of Córdoba. Córdoba, Spain.
- [5] Duque, M., y Néstor, D. (2008). *Student Model for Adaptive Systems of Virtual Education*. Medellín, Colombia
- [6] Mladen, W. (2008). *Data Mining y Data Warehousing*