



UNIVERSIDAD TÉCNICA PARTICULAR DE LOJA
La Universidad Católica de Loja

ÁREA TÉCNICA

**TÍTULO DE INGENIERO EN SISTEMAS INFORMÁTICOS Y
COMPUTACIÓN**

Análisis de patrones de comportamiento de los estudiantes que utilizan dispositivos móviles en su proceso de enseñanza aprendizaje, a través de técnicas de minería de datos.

TRABAJO DE TITULACIÓN

AUTORA: Samaniego Franco, Jennifer Betzabé

DIRECTORA: Agila, Martha Vanessa, Mgs.

**LOJA - ECUADOR
2016**



Esta versión digital, ha sido acreditada bajo la licencia Creative Commons 4.0, CC BY-NY-SA: Reconocimiento-No comercial-Compartir igual; la cual permite copiar, distribuir y comunicar públicamente la obra, mientras se reconozca la autoría original, no se utilice con fines comerciales y se permiten obras derivadas, siempre que mantenga la misma licencia al ser divulgada. <http://creativecommons.org/licenses/by-nc-sa/4.0/deed.es>

Septiembre, 2016

APROBACIÓN DE LA DIRECTORA DEL TRABAJO DE TITULACIÓN

Mgs.

Martha Vannesa Agíla Palacios

DOCENTE DE LA TITULACIÓN

De mi consideración:

El presente trabajo de titulación; “Análisis de patrones de comportamiento de los estudiantes que utilizan dispositivos móviles en su proceso de enseñanza aprendizaje, a través de técnicas de minería de datos” realizado por Jennifer Betzabe Samaniego Franco, ha sido orientado y revisado durante su ejecución, por cuanto se aprueba la presentación del mismo.

Loja, Mayo 2016

f).....

Mgs. Martha Vannesa Agíla Palacios

DECLARACIÓN DE AUTORÍA Y CESIÓN DE DERECHOS

“Yo, Jennifer Betzabe Samaniego Franco declaro ser autor (a) del presente trabajo de titulación: Análisis de patrones de comportamiento de los estudiantes que utilizan dispositivos móviles en su proceso de enseñanza aprendizaje, a través de técnicas de minería de datos, de la Titulación Sistemas Informáticos y Computación, siendo Mgs. Martha Vanessa Agila Palacios director (a) del presente trabajo; y eximo expresamente a la Universidad Técnica Particular de Loja y a sus representantes legales de posibles reclamos o acciones legales. Además certifico que las ideas, conceptos, procedimientos y resultados vertidos en el presente trabajo investigativo, son de mi exclusiva responsabilidad.

Adicionalmente declaro conocer y aceptar la disposición del Art. 88 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 o trabajos de titulación que se realicen con el apoyo financiero, académico o institucional (operativo) de la Universidad”

f.

Jennifer Betzabe Samaniego Franco

1104678899

DEDICATORIA

Dedico este trabajo especialmente a mi madre quién día a día lucho por darme un mejor futuro, quien con su ejemplo me enseñó que a pesar de las adversidades de la vida todo se lo puede de la mano de Dios.

A mi padre que a pesar de la circunstancias de la vida siempre ha estado a nuestro lado brindándonos su apoyo.

A mi querido Leonardo, quien siempre me brindó su apoyo en los buenos y malos momentos con sus bromas, consejos dándome las fuerzas que necesitaba para continuar.

A mi pequeño hijo que aún no lo conozco pero que estos dos meses ha sido mi mayor inspiración para cada paso que doy en mi vida.

AGRADECIMIENTOS

Para alcanzar algo en la vida siempre hay varias personas que nos extienden su mano, amistad y nos ayudan a conseguir nuestras metas es por esto que quiero aprovechar este espacio para poder expresarles mis agradecimientos.

Agradezco primeramente a Dios porque es nuestro padre supremo del que siempre he sentido su protección, además de brindarme la sabiduría necesaria para poder culminar esta etapa de mi vida.

A toda mi familia: mis padres, hermana, tíos, abuelitos y primos que son mi pilar fundamental en mi vida gracias porque siempre me han brindado su apoyo incondicional y motivación para alcanzar mis sueños.

A mis compañeros de trabajo: Mayuri, Rodrigo, Luis, Alex, Pablito, Fabricio, Jorge, Marlon, Marcelo, Paola, Mayra e Inesita gracias por el apoyo que me brindaron en todo momento y además de compartir sus conocimientos y consejos tanto laborales como académicos.

Mis más sinceros agradecimiento a mi directora de tesis Marthita Agila, quien me ha guiado en todo el desarrollo de este proyecto; gracias por su disponibilidad de tiempo a pesar de la distancia, por su entrega e interés para que todo salga bien y más que nada gracias por cada consejo tanto académico como personal.

Al Ing. Juan Carlos Torres quien siempre me brindo su ayuda en cada momento que creía necesario para el desarrollo de este proyecto.

Finalmente agradecer a mis queridos amigos: Priscila, Kruskaya, Karina, Walter, Stalin, Jonathan con los que cuento siempre en momentos alegres y tristes. Al igual Henry, Cesar, Eduardo, Fernada, Fefis, Silvia, Gladys, Samuel por ser amigos y familia a la vez siempre brindándome su apoyo incondicional.

ÍNDICE DE CONTENIDO

DECLARACIÓN DE AUTORÍA Y CESIÓN DE DERECHOS.....	iii
DEDICATORIA.....	iv
AGRADECIMIENTOS	v
ÍNDICE DE TABLAS	ix
ÍNDICE DE FIGURAS	x
RESUMEN.....	1
ABSTRACT	2
INTRODUCCIÓN	3
CAPÍTULO I: MARCO TEORICO.....	4
1.1 E-books o libros electrónicos.....	5
1.1.1 Características.....	5
1.1.2 Formatos.....	6
1.1.3 Dispositivos.....	8
1.1.4 Tipos De Ebooks.....	8
1.1.5 Política De Adquisición.	9
1.1.6 Gestión De Derechos Digitales (DRM).....	10
1.1.7 Ventajas.....	11
1.1.8 Desventajas.	11
1.1.9 Ebooks En La Educación.	12
1.2 Tipos de aprendizaje con tecnología.	12
1.2.1 E-learning, B-learning y otros.....	13
1.2.2 M – Learning.....	14
1.3 Minería de datos.....	20
1.3.1 Tareas Y Métodos.....	22
1.3.2 Herramientas.	24
1.4 Metodología para proyecto de minería de datos.....	29
1.4.1 Comparación de metodologías	29
1.4.2 Crisp – Dm (Chapman, y otros, 2000).....	30
1.5 Proyectos relacionados	35
1.5.1 Patrones de M – Learning en el aula virtual. (López Hernández & Silva Pérez, 2014)	35

1.5.2	Uso de Ebooks por los alumnos de la universidad de Andrews: un estudio de las actitudes, percepciones y comportamientos (Marques de Oliveira, 2012).....	36
1.5.3	El uso del libro electrónico en una biblioteca universitaria para docentes, alumnos de pregrado y postgrado: actitudes y comportamientos de los usuarios (Allen Shelburne, 2013)	37
CAPÍTULO II: ANÁLISIS DE LA PROBLEMÁTICA Y DISEÑO DE LA SOLUCIÓN.....		38
2.1	Análisis de la problemática	39
2.2	Diseño de la solución	40
2.2.1	Variables.....	42
2.2.2	Herramienta de minería de datos	44
2.2.3	Técnicas de minería de datos	44
CAPÍTULO III: Proceso de minería de datos		47
3.1	Fase I. compresión del negocio.....	48
3.1.1	Determinación de los Objetivos del Negocio	48
3.1.2	Evaluación de la Situación	48
3.1.8	Determinación de los Objetivos de la Minería	51
3.1.3	Plan del Proyecto.....	51
3.2	Fase II. Comprensión de los datos	52
3.2.1	Muestreo.....	53
3.2.1	Recolección de Datos	55
3.2.2	Descripción de los datos	56
3.2.3	Exploración de los datos	62
3.2.4	Verificación de los datos	64
3.3	Fase III. Preparación de datos.....	65
3.3.1	Selección de datos.....	65
3.3.2	Limpieza de datos.....	67
3.3.3	Construcción e Integración de Datos	68
3.4	Fase IV. Modelado	69
3.4.1	Selección de la técnica del modelado	69
3.4.2	Generación de la prueba de diseño	69
3.4.3	Construcción del modelo: Entorno Virtual de Aprendizaje (EVA)	70

3.4.4 Construcción del modelo: interacción en la Tableta digital mediante SPSS	80
3.4.4 Evaluación del modelado	86
CONCLUSIONES.....	89
RECOMENDACIONES.....	91
BIBLIOGRAFÍA.....	92
ANEXOS	98
Anexo 1: Tablas del entorno virtual de aprendizaje (eva)	99
Anexo 1 – A: mdl_user.....	99
Anexo 1 – B: mdl_log.....	99
Anexo 1 – C: Tablas del EVA Relacionadas	100
Anexo 1 – D: Variables de modules - acciones.....	100
Anexo 2: Tabla de datos demográficos (syllabus)	101
Anexo 2 – A: datos_personales	101
Anexo 3: Vistas servidor gestión ebooks (GRAMMATA)	101
Anexo 3 – A: vistas	101
Anexo 4: Consultas sql.....	102
Anexo 4 – A: Consultas EVA	102
Anexo 4 – B: Consultas gestión Ebooks	102
Anexo 5: Distribución por centros.....	102
Anexo 6: APLICACIONES.....	105
Anexo 6 – A: Aplicaciones que conforman la categorización de las variables de aplicaciones en la Tableta digital	105
Anexo 6 – B: Juegos que conforman la categorización de las variables de juegos en la Tableta digital	107
Anexo 7: FRECUENCIAS DE INTERACCIÓN	110
Anexo 7 – A: Género, estado_civil, edad EVA	110
Anexo 7 – B: Género, estado_civil, edad Tableta digital	111

ÍNDICE DE TABLAS

Tabla 1 <i>Formato Ebooks</i>	7
Tabla 2 Evolución de la Sociedad.....	14
Tabla 3 Proyectos M –Learning América Latina	18
Tabla 4 Correspondencia entre tareas y técnicas	24
Tabla 5 Comparación de Herramientas	28
Tabla 6 Comparación de Metodologías.....	29
Tabla 7 Recursos Personales	49
Tabla 8 Recursos de Información.....	50
Tabla 9 Software a Utilizar.....	50
Tabla 10 Plan del Proyecto.....	51
Tabla 11 Variables a utilizar mdl_user.....	57
Tabla 12 Variables a utilizar mdl_log.....	58
Tabla 13 Datos personales en la tabla Syllabus	59
Tabla 14 Vista biusuario	60
Tabla 15 Vista bidispositivo	61
Tabla 16 Vista bidispositivousuario	61
Tabla 17 Vista bihistoricoaplicacion	61
Tabla 18 Vista biaplicacion	61
Tabla 19 Categorización edades.....	62
Tabla 20 Distribución estado civil	63
Tabla 21 Distribución género.....	63
Tabla 22 Interacción en el EVA	65
Tabla 23 Evaluaciones en el Eva	65
Tabla 24 Blog en el EVA	66
Tabla 25 Variables no usadas para la interacción en el EVA.....	66
Tabla 26 Variables interacción Ebooks	67
Tabla 27 Análisis estadístico descriptivo de frecuencias.....	71
Tabla 28 Categorización de la interacción en el EVA.....	78
Tabla 29 Distribución de patrones de comportamiento	78
Tabla 30 Categorización de la interacción en la Tableta digital digital	83
Tabla 31 Interacción Tableta digital digital - EVA.....	85
Tabla 32 Prueba de verosimilitud EVA.....	86
Tabla 33 Bondad de ajuste EVA.....	87
Tabla 34 Pseudo R-cuadrado.....	87
Tabla 35 Prueba de verosimilitud Tableta digital digital	87
Tabla 36 Bondad de ajuste Tableta digital digital.....	87
Tabla 37 Pseudo R-cuadrado Tableta digital	88

ÍNDICE DE FIGURAS

Figura 1. Formatos Ebooks.....	6
Figura 2 Modelo del DRM en un libro electrónico.	10
Figura 3 Ventajas Ebooks	11
Figura 4 Evolución de la Sociedad	15
Figura 5 Definición M - Learning.....	16
Figura 6 Ecosistema de M – Learning de la región NOA.....	17
Figura 7 Proyecto MoMath	19
Figura 8 Esquema del Proceso KDD	21
Figura 9 Fases del proceso de extracción del conocimiento	22
Figura 10 Encuesta Metodología de Minería de Datos	30
Figura 11 Modelo Crisp - DM.....	31
Figura 12 Ciclo de Vida Crisp - DM.....	31
Figura 13 Fases Crisp - DM	32
Figura 14 Compresión del Negocio	32
Figura 15 Compresión de Datos	33
Figura 16 Preparación de Datos	33
Figura 17 Modelado	34
Figura 18 Evaluación	35
Figura 19 Ecosistema M –Learning UTPL	39
Figura 20 Solución.....	42
Figura 21 Variables de Datos Personales Modelo Descriptivo.....	42
Figura 22 Variables de Interacción en el EVA.....	43
Figura 23 Variables de Interacción en las Tableta digital	44
Figura 24 Error Cuadrático.....	45
Figura 25 Partes de un Árbol J48	46
Figura 26 Afluencia de estudiantes del proyecto Ebooks	53
Figura 27 Número de Tablet Entregadas a los Estudiantes.....	54
Figura 28 Fórmula para la obtención del tamaño de la muestral	54
Figura 29 Interacción en el Entorno Virtual (EVA).....	55
Figura 30 Variables de Datos Personales	56
Figura 31 Interacción en el Servidor de Gestión Ebooks.....	56
Figura 32 Gráfica de distribución por edades	62
Figura 33 Gráfica de distribución del estado civil	63
Figura 34 Gráfica de distribución por género.....	64
Figura 35 Distribución por centros.....	64
Figura 36 Conjunto de datos interacción EVA.....	68
Figura 37 Conjunto de datos Interacción en la Tableta digital	69
Figura 38 Valores atípicos EVA	72
Figura 39 Tabla de valores KMO.....	73
Figura 40 Determinante de la matriz de correlaciones EVA	73
Figura 41 Pruebas KMO y Bartlett EVA	74
Figura 42 Varianza EVA.....	74
Figura 43 Clusterización numCluster 2	75
Figura 44 Clusterización numCluster 3	76
Figura 45 Clusterización numCluster 4	76
Figura 46 Clusterización interacción EVA SPSS	77
Figura 47 Asignación del número de cluster	77
Figura 48 Patrones de comportamiento en el EVA.....	78
Figura 49 Porcentaje interacción en el EVA.....	79

Figura 50	Uso de aplicaciones en la Tableta digital digital.....	80
Figura 51	Uso de juegos en la Tableta digital digital	80
Figura 53	Pruebas KMO y Bartlett Tableta digital digital	81
Figura 54	Varianza Tableta digital	82
Figura 55	Clusterización interacción Tableta digital digital SPSS	82
Figura 56	Registro de interacción en las Tableta digital digital	83
Figura 57	Patrones de comportamiento en las Tableta digital	84
Figura 58	Porcentaje interacción Tableta digital digital.....	85

RESUMEN

En el presente trabajo se obtuvo un modelo de minería de datos basándose en la metodología Crisp – DM, mediante la información de los estudiantes de primer ciclo de la modalidad abierta y a distancia perteneciente al proyecto Ebooks se analizó su interacción tanto en el Entorno Virtual de Aprendizaje (EVA) y en la Tableta digital, se obtuvieron sus patrones de comportamientos, para poder conocer los perfiles que tienen los estudiantes que utilizan la Tableta digital en su aprendizaje. El presente estudio le permitirá a la universidad conocer si los estudiantes usan la tableta netamente para su formación académica o para otros fines permitiendo analizar la implementación de nuevas herramientas dentro de la Tableta digital.

Palabras Claves: patrones de comportamiento, minería de datos, metodología Crisp – DM, m – learning, Ebooks, clúster.

ABSTRACT

In this paper a model of data mining based on the Crisp methodology was obtained - DM, by information undergraduate students of open and distance modality belonging to Ebooks project interaction was analyzed both the Virtual Learning Environment (EVA) and digital Tablet, behavior patterns were obtained in order to know the profiles that have students using digital tablet in their learning. This study will allow the university to know if students use tablet purely for their education or for other purposes it possible to analyze the implementation of new tools within the digital tablet.

KEYWORDS: behavior patterns, data mining methodology Crisp - DM, m - learning, Ebooks, cluster.

INTRODUCCIÓN

En la actualidad, la tecnología en general es el eje principal para el desarrollo de nuestras actividades cotidianas, aún más la tecnología móvil que está marcando una mayor diferencia ya que se ha vuelto indispensable. Es así como en la educación, la tecnología móvil está cambiando parte del modelo educativo con una de sus innovaciones como los ebooks (libros electrónicos), los cuales brindan al estudiante una manera interactiva en el proceso de aprendizaje. Muchas Universidades están implantando esta tecnología, como bibliotecas digitales que en si es un catálogo el cual permite acceder a la información de la biblioteca desde cualquier lugar, la diferencia que marca el proyecto EBOOKS de la Universidad Técnica Particular De Loja es que su objetivo principal es la entrega de todo el material bibliográfico en formato digital y en perpetuidad, brindándoles la ventaja de contar con su material en todo momento y la oportunidad de aprovechar al máximo su tiempo disponible para realizar su tareas.

Si bien existen estudios relacionados al aprendizaje con tabletas, éstos son enfocados a educación primaria y secundaria y a contextos de consulta en bibliotecas o en el aula presencial, aspectos muy diferentes a nuestra realidad, que es un contexto de educación superior, con modalidad de estudios a distancia.

Con la obtención de los patrones de comportamiento se conocerán las estrategias de estudio de cada estudiante además de obtener los grupos de comportamiento de los mismos, por lo tanto ésta será la base que determinará si se necesita la implementación de nuevas herramientas en la Tableta digital.

El presente estudio se planteó mediante los siguientes objetivos específicos:

- Analizar las experiencias de universidades que han implementado y obtenido resultados con tecnología m – learning.
- Recolectar, explorar y verificar datos de los estudiantes de la Modalidad a Distancia de la UTPL del Proyecto Ebook .
- Seleccionar, limpiar, integrar y formatear los datos para determinar grupos de comportamiento de los estudiantes de la Modalidad a Distancia de la UTPL del Proyecto Ebook.
- Definir el modelado, diseñar el plan de pruebas y validar el modelado.

La metodología Crisp –DM fue utilizada luego de haber realizado una comparación de metodologías y concluir que era la indicada para el desarrollo del proyectos, las herramientas que se utilizaron fueron WEKA y SPSS en el desarrollo del modelo se explica por qué se usó ambas herramientas.

CAPÍTULO I: MARCO TEORICO

1.1 E-books o libros electrónicos

Una de las definiciones más representativa es la otorgada por (Oxford Dictionaries, 2013), en donde define a un Ebook como *“una versión electrónica de un libro impreso que se puede leer en un ordenador personal o dispositivo de mano diseñado específicamente para este propósito”*.

Pero a su vez se puede definir en función de cuatro factores que integran un Ebook como lo menciona (Galisteo del Valle, Fernández, & Garrido Sergio, 2011):

- Soporte - formato.- Cada formato tienen una norma diferente al igual que el soporte es diferente dependiendo a que empresa pertenecen.
- Tipo de información.- Existe diferente tipo de información para los Ebooks ya sea en texto plano o pueden ser Ebooks interactivos con imágenes, sonidos o video.
- Organización de la información.- Lineal o en árbol con opciones de navegación.
- Tipos de interacciones con la información.- Las interacciones que se pueden realizar en el lector del libro digital son: notas, búsquedas, audio, zoom, otras interacciones (dependiendo de las opciones que brinde el lector).

1.1.1 Características.

Un libro electrónico debe seguir el estándar planteado del libro físico, en cuanto a la calidad de la información y su organización, lo que permite brindar accesibilidad y soporte para actividades formativas que refuercen el aprendizaje. La diferencia con el libro físico es que mediante ejercicios propuestos, se fomenta a los estudiantes el empleo de sus competencias cognitivas, para el desarrollo de la solución de problemas específicos (Aedo & Díaz, 2000).

Un estudio realizado sobre la integración de las TIC (Tecnología de la Información de la Comunicación) con los libros digitales en la educación, desarrollada por el sitio web educativo “Aula Planeta”, quienes realizaron un sondeo a docentes de España, quienes manifestaron que el libro de texto digital debe ser como lo menciona (Chuquiyauri, 2014), teniendo en cuenta las siguientes características:

- Interactivo.
- Actualizable.
- Audiovisual.
- Personalizable.
- Adaptable.

Además, sugieren que los libros electrónicos tengan integradas características como lo menciona (De Oliveira, Camacho, & Gisbert , 2014):

- Búsquedas
- Referencias Cruzadas.
- Enlaces a hipertextos
- Marcas
- Anotaciones
- Subrayado
- Objetos multimedia

1.1.2 Formatos.

Al momento de hablar de Ebooks hay muchos como los menciona (Baratz, 2014) en la (Figura 1:1):

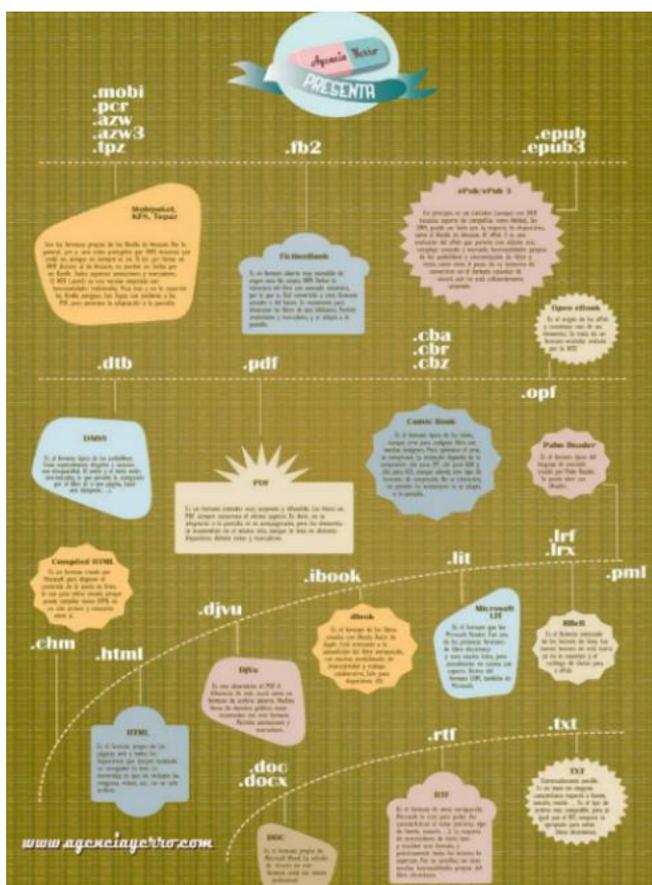


Figura 1. Formatos Ebooks

Fuente: (Baratz, 2014).

Pero los formatos más usados según (Universia, 2013) se observa en la siguiente tabla:

Tabla 1 Formato Ebooks

FORMATO	DESCRIPCIÓN
Mobi	<p>Formato nativo del lector de Amazon Kindle</p> <p>Basado en XHTML</p> <p>No permite que un dispositivo lector use un DRM distinto al ofrecido por MobiPocket.</p>
Epub	<p>Sucesor del Open eBook</p> <p>Basado en XML</p> <p>Admite desde imágenes a gráficos vectoriales y es el candidato más firme para ser considerado como el MP3 de los libros.</p> <p>Ligero peso de sus archivos además disminuye los tiempos de carga y permite una mayor manejabilidad.</p> <p>Formato totalmente abierto y diseñado expresamente para la lectura digital.</p> <p>Está compuesto por tres estándares: OPS, OPF y OCF.</p> <p>OPS: Open Publication Structure</p> <p>OPF: Open Packaging Format</p> <p>OCF: Open Container Format</p>
Txt	<p>No es un formato como tal o más bien que es texto sin formato, llamado texto plano.</p> <p>Probablemente el tipo de fichero más compatible que pueda existir, ya que carece precisamente de formato.</p> <p>Puede almacenar grandes cantidades de texto en muy poco espacio.</p> <p>Carece de imágenes, tablas, vínculos o cualquier elemento visual que no sean palabras, números o símbolos de puntuación.</p>
Pdf	<p>Estándar abierto y uno de los más extendidos.</p> <p>Formato estándar, con mucha difusión</p> <p>Debido a que no ha sido diseñado para modificar la composición de las páginas, a veces, no es posible adaptarlo al tamaño de pantalla de los eReaders de forma correcta.</p>
Html	<p>Es el mismo formato de las páginas web</p> <p>Con este formato se podrá ver en equipo en el navegador web.</p> <p>Totalmente estándar y prácticamente todos los lectores lo reconocen.</p> <p>Posibilidad de incorporar video y audio</p> <p>Solo se utiliza en revistas u otras obras que posean pocas páginas</p>

Fuente: Elaboración Propia

1.1.3 Dispositivos.

Las Tableta digital y los dispositivos eReader sirven para leer Ebooks, pero existen características que los diferencia significativamente, como las que se presenta a continuación (Comunidad Movistar, 2012):

1.1.3.1 *Tableta digital.*

- **Pantallas LCD:** brindan la ventaja de leer en la oscuridad por la luz propia de la pantalla, pero se debe tener en cuenta, que leer por varias horas continuas resulta cansado para la vista.
- **Tamaño:** generalmente son más grande que un dispositivo eReaders.
- **Precio:** existe gran diversidad de precios en la Tableta digital pero siempre serán más económicas que los dispositivos eReaders.
- **Opciones Lector:** depende de la Aplicación que se elija para acceder a un Ebook, cada aplicación soporta diferentes formatos y en la actualidad todos los lectores brindan varias opciones interactivas.
- **Usabilidad:** opciones como navegación, juegos, música, cámara, etc.
- **Batería:** necesita ser cargada al menos una vez al día.

1.1.3.2 *E-Readers.*

- **Tinta electrónica:** su mayor ventaja es su pantalla con tinta electrónica, las cuales no afectan a las vista, teniendo en cuenta que no se pueden visualizar imágenes solo texto en color negro.
- **Tamaño:** cada vez son más pequeños, lo que les permite ser portables.
- **Precio:** son accesibles debido a su bajo costo, además los Ebooks tienen un precio muy parecido al libro físico.
- **Opciones:** limitado, utiliza marcadores para resaltar partes importantes en el documento y permite cambiar el tipo de letra.
- **Usabilidad:** aun muchos poseen teclado.
- **Batería:** dura de 4 a 5 días.

1.1.4 Tipos De Ebooks.

Los libros electrónicos se clasifican dependiendo del tipo de información que contiene, las funciones que ejecutan según (Cabrera Frías & Malágon Hernández, 2009) son:

- **Libro de Texto:** texto plano.
- **Libro Parlante:** combina texto plano y en audio.
- **Libro de Imagen Estática:** posee solamente imágenes.
- **Libro de Imagen en Movimiento:** posee animación y vídeo.

- **Libro Multimedia:** combina imágenes, animación, audio y video.
- **Libro Polimedia:** posee la ventaja de transferir información de diversas formas: por papel, cd, internet, redes locales, etc.
- **Libro Hipermedia:** parecidos a los libros multimedia, con la diferencia que no siguen una organización estricta de su información.
- **Libro Electrónico Inteligente:** brindar aprendizaje a su usuario y aprende de ellos empleando inteligencia artificial.
- **Libro Telemedia:** posee características innovadoras, como la posibilidad de utilizar recursos como: correo electrónico, bibliotecas digitales, videoconferencias.

1.1.5 Política De Adquisición.

Para la obtención de los Ebooks existen algunos modelos de adquisición como lo mencionan (Cordón García & Arévalo, 2010), (Universidad Complutense, 2011) :

- Modelo de adquisición.
- Uso.
- Tipo de proveedor.
- Modo de selección.

1.1.5.1 Modelo de adquisición: Compra frente a suscripción.

Compra a perpetuidad: la institución adquiere el contenido de forma indefinida.

Suscripción: la institución adquiere derecho sobre el material seleccionado y sus actualizaciones, durante un periodo de tiempo y bajo el pago de una cuota anual.

1.1.5.2 Uso: uso único o colectivo.

Un aspecto importante es el número de usuarios, existen empresas que dependiendo del contenido contratado permiten el ingreso ilimitado de usuarios, de acuerdo a sus políticas.

Tipo de proveedor: agregador o editor: El agregador es una plataforma común a varias editoriales, o bien directamente desde la web del propio editor del libro. Un ejemplo de agregador es Amazon, Libranda, Luarna, etc.

Modo de selección: “Pick and choose” versus Colección.

“Pick and choose” (del Español “Escoge y Elige”) es la elección de un libro individualmente título a título, tal como se hacía habitualmente en el modelo tradicional en formato impreso, frente al modelo de suscripción a colecciones seleccionadas por el propio editor.

1.1.6 Gestión De Derechos Digitales (DRM).

Como lo menciona (Groenenboom & Helberger, 2006) los sistemas de gestión de derechos digitales o más conocido como DRM (Digital Rights Management), se aplican para proteger y vigilar el uso adecuado de la información. Existen varios sistemas DRM utilizados en medios físicos como digitales.

Los sistemas de gestión de derechos digitales están contenidos en los libros electrónicos, mediante el equipo en el que los visualizan, por tal motivo, están estrechamente relacionados con el hardware y software para que se pueda utilizar DRM. La información editorial del Ebook queda almacenada con un número de identificación de usuario para relacionar la descarga con el usuario y el equipo (Alonso, 2013).

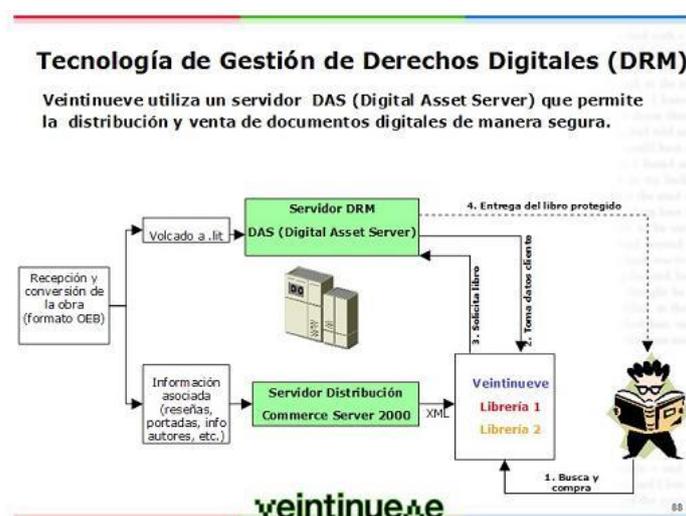


Figura 2 Modelo del DRM en un libro electrónico.

Fuente: (Ramos, 2004)

Los cuatro programas más expendidos en el mercado son: Adobe Digital Editions, Amazon DRM, Apple DRM y Barnes and Noble DRM (Alonso, 2013).

Los DRM emplean diversas técnicas para monitorear y controlar el uso del contenido digital, principalmente como lo menciona (Groenenboom & Helberger, 2006):

- **Encriptamiento (encryption).**- es la codificación de la clave que restringe el acceso al contenido que se encuentra en el hardware o software que se utiliza como lector.
- **Marcado (marking).**- mediante el marcador se brinda información de las características del contenido como: protección contra copia, autor titular de los derechos, tipos de usos autorizados esto lo realizar el autor antes de su venta.
- **Incompatibilidad selectiva (incompatibility selective).**- es la compatibilidad existente con algunos dispositivos y la incompatibilidad con otros, esto depende del formato establecido del Ebook, además del tipo de hardware o software que se utilice.

1.1.7 Ventajas.

Según (Arévalo, Córdon, & Gómez, 2011), menciona las siguientes ventajas y desventajas:

1. Acceso en todo tiempo y lugar.
2. Espacio.
3. Buscabilidad.
4. Texto enriquecido.
5. Portabilidad y Movilidad.
6. Conservación del medio ambiente.
7. Durabilidad.

Además (López Herrero, 2010), menciona las siguientes características:

- Navegabilidad.
- Actualización.
- Educación inclusiva.

Cabe recalcar el por qué se habla de educación inclusiva, al llegar con la tecnología a niños, adolescentes y adultos, sin importa su condición económica, social y cultural brindando a si una igualdad de condición brindando una educación justa y equitativa.

En la Figura 3, se resume las ventajas de tener un libro electrónico:



Figura 3 Ventajas Ebooks

Fuente: (Chuquiyauri, 2014).

1.1.8 Desventajas.

Las principales desventajas de los Ebooks son:

1. Necesidad de dispositivos de lectura.
2. Fragilidad.

3. Variedad de formatos.
4. Problemas de intervención editorial.
5. Conexión.

1.1.9 Ebooks En La Educación.

La tecnología por sí sola no innova la enseñanza ni el aprendizaje, al momento de combinarla con los Ebooks y la educación, se alcanzan mejores resultados en el aprendizaje, además se debe tener un repositorio de actividades y contenidos innovadores desarrollados con los Ebooks y dispositivos móviles, para aumentar el interés en los alumnos. (Salamanca & Sánchez, 2010).

Desde otro punto de vista, el cambio educativo es mucho más complejo que incorporar Tableta digital o dispositivos de lectura en el aula y ofrecer contenidos digitales en la nube.

“La innovación pedagógica está en la transformación de las creencias y prácticas de los agentes educativos (lo cual es lento y complicado), más que en el cambio de los productos tecnológicos” (Area, 2013).

Para (López Herrero, 2010), el uso de Ebooks en la educación ha permitido un acceso no lineal a la información.

Un estudio desarrollado por (De Oliveira, Camacho, & Gisbert , 2014) a un grupo de estudiantes de primaria, concluye que el uso de Ebooks en la educación, permite a los alumnos convertirse en protagonistas, mediante el trabajo individual y colaborativo. Siendo el profesor quien orienta acerca de dichos trabajos.

1.2 Tipos de aprendizaje con tecnología.

Para entender de una mejor manera la clasificación de los tipos de aprendizaje con la tecnología, se debe tener en cuenta el significado de un Sistema de Gestión de Aprendizaje LMS.

LMS (del inglés Learning Management System) es un sistema informático utilizado para administrar, distribuir y controlar actividades de formación en una modalidad de estudio a distancia. (NewWweb, 2014).

Para el (Instituto de Tecnologías Educativas, n.d.), los LMS deben incluir las siguientes herramientas:

- Gestión y administración de usuarios.
- Gestión de cursos, grupos o actividades.

- Gestión de las herramientas de comunicación síncronas y asincrónicas, correo electrónico interno, foros, chat o web chat, blogs, wikis, videoconferencia, tableros de anuncios y avisos, etc.

Un sistemas de gestión de aprendizaje es una de las herramientas más utilizadas en los modelos de aprendizaje, por ello un LMS es más conocido como una plataforma de aprendizaje (García Peñalvo, 2005).

Al hablar de los tipos de aprendizaje, se tiene en cuenta que existen muchos autores que nombran diferentes tipos de aprendizaje según la tecnología, como se menciona a continuación:

(Padilla, 2014) Nombra a los tipos de aprendizaje como los nuevos modelos de aprendizaje.

En contraste (Castro, 2014) habla en un seminario acerca de “Las Nuevas Formas de Enseñanza en la Universidad Digital” el cual menciona esta clasificación como una de las tendencias en el uso de la tecnología para la educación.

En conclusión muchos autores mencionan la misma clasificación con un nombre diferente.

A continuación se menciona los tipos de modelos de aprendizaje (Padilla, 2014; Castro, 2014):

1.2.1 E-learning, B-learning y otros

- **E-learning.-** tipo de aprendizaje obtenido a través del uso del internet, con el uso de diferentes aplicaciones informáticas, convirtiéndose en uno de los aprendizajes más usados (E-abclearning, 2011).
- **B-learning.-** este aprendizaje combina la educación a distancia y presencial (e-learning) (Ciberaula, 2012).
- **U-learning.-** es el aprendizaje que se genera desde cualquier lugar y en cualquier momento el mismo que no depende de un dispositivo móvil específico sino que puede obtener conocimiento mediante cualquier medio tecnológico que pueda obtener y manipular información (Herrera Cuenca, 2013).
- **P-learning.-** se podría considerar una evolución del M-Learning, utiliza ordenadores que obtienen información sobre el contexto de aprendizaje mediante pequeños dispositivos integrados como sensores o etiquetas (Villar , 2011).
- **C-learning.-** se trata de cualquier tipo de aprendizaje obtenido usando medios sociales, virtuales que permiten un trabajo comunitario además de espacios para la comunicación y colaboración (Martí, 2012).

- **T-learning.-** es un sistema digital que utiliza la interactividad y personalización, es más conocido como aprendizaje televisivo (Virtual, Aula, 2012).
- **M-learning.-** este aprendizaje se lo obtiene mediante dispositivos móviles (Edukanda, s.f.).

1.2.2 M – Learning.

Para entender el significado de M-Learning, es necesario recordar el avance de la sociedad, en especial en la cultura y educación en relación con la evolución científica y tecnológica actual.

Como explica (Moravec & Cobo, 2011) la evolución de la sociedad se ha visto marcada por los cambios y cómo estos cambios han influido en los paradigmas sociales. Hace una revisión cronológica, basada en las transformaciones tecnológicas y económicas.

Se observa en la Tabla 2, las transformaciones que cita (Moravec & Cobo, 2011) al igual que (Quispe Ortega, 2011):

Tabla 2 Evolución de la Sociedad

EVOLUCIÓN DE LA SOCIEDAD	
Web 1.0	Personas conectándose a la Web y la Web como punto de información estática.
Web 2.0	Personas conectándose a personas, la inteligencia colectiva como centro de información y la Web es sintáctica.
Web 3.0	Aplicaciones Web conectándose a aplicaciones Web, las personas siguen siendo el centro de la información y la Web es semántica.
Web 4.0	Personas conectándose con Personas y aplicaciones Web de forma ubicua, se añaden tecnologías como la Inteligencia Artificial, la Voz como vehículo de intercomunicación para formar una Web Total.

Fuente: (Moravec & Cobo, 2011) & (Quispe Ortega, 2011)

Las sociedades 2.0 y 3.0 representan el presente y el futuro inmediato, están caracterizadas por el mayor uso de internet, en donde se encuentra información además de producirla y sobre todo, compartirla para convertirla finalmente en un conocimiento colectivo.

En la Figura 4, se observa, de una manera más clara el proceso de evolución de la sociedad:

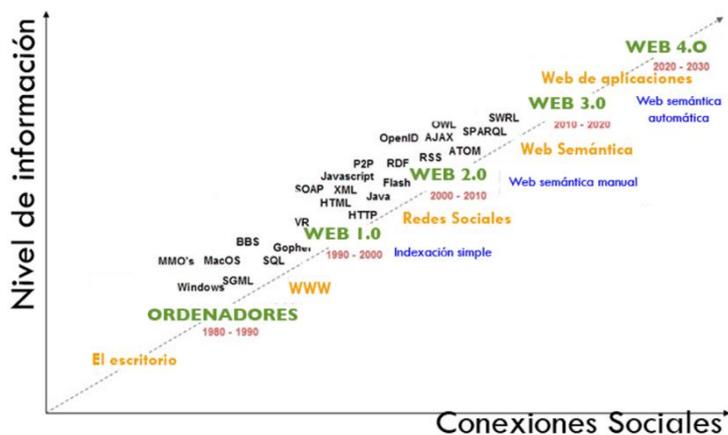


Figura 4 Evolución de la Sociedad

Fuente: (Tapia, 2013)

Luego de haber recordado la evolución de la sociedad se hablara de lo que es M-Learning.

M-learning es el aprendizaje e-learning que se apoya en los dispositivos móviles y en la transmisión de información por medio de la red Wireless, lo que permite al profesor y al alumno mantener un contacto constante a cualquier momento del día, fomentando con ello una educación personalizada y adaptándose a las necesidades de los alumnos en cada momento. (Ramírez Montoya, 2008) (Moreno Guerrero , 2011).

El aprendizaje M-learning o también conocido como aprendizaje móvil. (Traxle, 2007) Define el aprendizaje móvil como parte de una nueva concepción de la sociedad siendo la tecnología dominante y principalmente los dispositivos portables o de bolsillo.

Aplicar M-learning incentiva a usar métodos innovadores de enseñanza eliminando aquellos conceptos anticuados y rígidos.

A diferencia de lo que menciona (Mohamed , El-Hussein , & Cronje) el cual define al M-learning dividiéndolo en tres áreas importantes:

- Movilidad de la tecnología.
- Movilidad del alumno.
- Movilidad de aprendizaje.

Esta división tripartita de la movilidad es común en la literatura actual sobre el tema y ha sido confirmado por los diseñadores que han utilizado la tecnología móvil para propósitos educativos. La Figura 1:5 es una representación gráfica de las tres divisiones para definir M-learning, que puede ofrecer un mayor nivel de instrucción educativa. En la práctica, la tecnología, el aprendizaje y el aprendiz, operan en un proceso continuo e ininterrumpido en el contexto social de la educación. Estos tres elementos son interdependientes y son igualmente importantes para convertir dispositivos móviles en instrumentos viables y útiles

para la entrega de contenidos de enseñanza en la educación superior (Mohamed , El-Hussein , & Cronje).

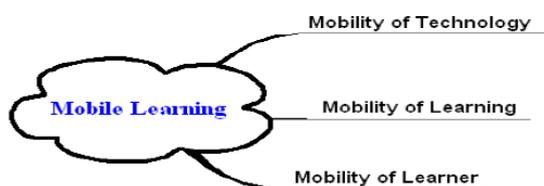


Figura 5 Definición M - Learning

Fuente: (Mohamed , El-Hussein , & Cronje)

(Roura Redondo, 2012) Relaciona las comunicaciones móviles además de las capacidades de comunicación electrónica. Al usar estas tecnologías se incentiva al estudiante aprender sin necesidad de estar en un lugar determinado, estas características brindan un avance hacia el aprendizaje ubicuo.

Factores que influyen en la generalización del uso de tecnologías móviles son:

- Expansión de la telefonía móvil.
- Aumento del uso de dispositivos móviles.
- La mejora en infraestructuras de redes inalámbricas.
- Generalización en el acceso móvil de banda ancha a Internet.
- Disminución del coste de los servicios relacionados con tecnologías móviles.
- Nuevos entornos de adquisición del conocimiento.
- Adecuación de contenidos para M-learning.

Las características tecnológicas asociadas al M-learning son:

- **Portabilidad:** permite llevar los dispositivos a todo lugar debido a su tamaño y peso.
- **Inmediatez y conectividad:** mediante la conexión a internet y aplicaciones de comunicación con el docente.
- **Ubicuidad:** permite aprender en cualquier lugar y momento.
- **Adaptabilidad de servicios:** aplicaciones adecuadas para el aprendizaje que el estudiante requiere desarrollar.

(Herrera, Fennema, & Sanz, 2010), interpreta el M-learning como un conjunto complejo de diversas formas de movilidad, tecnologías móviles, variedad de estudiantes, variedad de contenidos y contextos de aprendizaje.

Para este sistema complejo constituye según (Woodill, 2011), un ecosistema para formar un M-learning, el mismo que está formado por:

- Dispositivos.
- Infraestructura.
- Contextos.
- Contenidos.
- Plataformas.
- Herramientas.

Estos componentes son fundamentales en el marco del aprendizaje móvil (M-learning).

En la Figura 6, se observa el ecosistema de M-learning de NOA (Región Noroeste de Argentina):

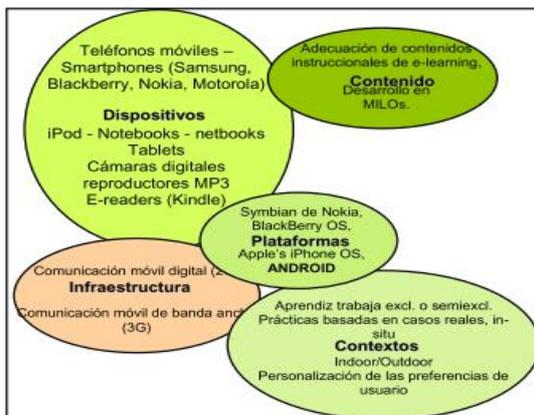


Figura 6 Ecosistema de M – Learning de la región NOA

Fuente: (Herrera, Fennema, & Sanz, 2010).

Basándose en el ecosistema antes mencionado, además de tener en cuenta el tipo de interacción que realizan se puede presentar el M-learning en tres modos diferentes como lo clasifica (Woodill, 2011) :

- **Modo recuperación de información (RI).**- acceso a la información, quiere decir que posee un único sentido de comunicación.
- **Modo recopilación y análisis de información (RAI).**- acceso a dos tipos de comunicación: tantos al acceso de información como a las respuestas o envíos de información.
- **Modo comunicación, interacción y colaboración en redes (CICR).**- interactúa con diversas comunidades de aprendizaje.

La revista International “Journal of Mobile y Blended Learning” es uno de varios indicadores de que el aprendizaje móvil en todo el mundo está creciendo a pasos agigantados y obteniendo una identidad propia (Woodill, 2011).

Los últimos seis o siete años se ha producido una gran cantidad de pilotos e iniciativas en todos los países, determinando que el M-learning toma el aprendizaje para los estudiantes, brindando ventajas para las comunidades y países en los que el acceso al aprendizaje era difícil, ahora es posible gracias a que el M-learning aumenta, enriquece y amplía la forma del aprendizaje.

Los factores ambientales han hecho que este desarrollo se produzca casualmente, enfrentando a la comunidad de aprendizaje móvil retos más amplios de escala, durabilidad, equidad, inclusión además de la pedagogía y tecnología (Traxler, Learning in a Mobile Age, 2009).

En la Tabla 3, se resume los proyectos M-learning en América Latina, los tres primeros son los más grandes proyectos:

Tabla 3 Proyectos M –Learning América Latina

Proyecto/Iniciativa	País	Descripción
Puentes Educativos	Chile	Los docentes utilizan teléfonos inteligentes para presentar videos educativos en clase.
Raíces de Aprendizaje Móvil	Colombia	Los docentes utilizan teléfonos inteligentes para presentar videos educativos en clase.
EMIA-SMILE	Argentina	Los estudiantes utilizan teléfonos inteligentes para participar en el aula en una actividad basada en la investigación.
Celumetrage	Argentina	Los estudiantes utilizan teléfonos móviles para grabar escenas de video de un corto.
Edumóvil	México	Los estudiantes utilizan programas de dispositivos móviles para aprender Matemática, castellano, Ciencias e Historia.
ViDHaC2	Chile	Los alumnos utilizan videojuegos educativos en teléfonos móviles para el aprendizaje auto dirigido en Ciencias dentro de la escuela.
Eduinnova	Chile	Los estudiantes usan netbooks para participar en una actividad colaborativa en el aula.
PocketSchool	El Salvador	Los niños utilizan dispositivos móviles para acceder a libros electrónicos y juegos educativos en el aprendizaje auto dirigido fuera de la escuela.
PSU Móvil	Chile	Los alumnos usan sus teléfonos móviles para acceder a ejercicios de práctica y a contenido para la Prueba de Selección Universitaria.
Evaluación de Aprendizajes a través de Celulares	Paraguay	Los estudiantes toman una prueba nacional por medio de sus teléfonos móviles en la escuela.
BlueGénesis	Colombia	Los universitarios utilizan teléfonos móviles con Bluetooth para intercambiar información con sus profesores y con sus pares.
M-iLab	México	Los universitarios usan iPhones para aprender Física.

Pos título de Especialización Superior en Educación a Distancia	Argentina	Los universitarios utilizan dispositivos BlackBerry para acceder a contenido y actividades de un curso semi presencial.
Proyecto Aprendizaje Móvil en el ITESM	México	Los universitarios utilizan dispositivos BlackBerry para acceder a contenido y actividades de un curso.
Proyecto Facebook	Argentina	Los estudiantes universitarios utilizan teléfonos móviles para grabar escenas de video de un corto.
Blackboard Mobile Learn+	México	Los universitarios utilizan iPhones y otros dispositivos de Apple con una versión móvil de Blackboard LMS para acceder a los contenidos y actividades de un curso.
Kantoo	Varios Países	El público general puede usar teléfonos móviles para acceder a lecciones de inglés (servicio pagado).
Programa Nacional de Alfabetización	Colombia	Jóvenes y adultos analfabetos utilizan un teléfono móvil con una tarjeta SIM especial para tomar un curso público de alfabetización.
PreveMóvil	Honduras	Los jóvenes reciben mensajes de texto en sus teléfonos móviles con información y consejos sobre cómo evitar el contagio con VIH.
Educación Móvil Continua en la Salud	Perú	Los trabajadores de la salud reciben mensajes de texto en sus teléfonos móviles para actualizar sus conocimientos profesionales.
DatAgro	Chile	Pequeños agricultores reciben mensajes de texto en sus teléfonos móviles sobre el mercado agrícola, el pronóstico del tiempo y noticias.

Fuente: (Unesco, 2012)

Estos son los proyectos más relevantes en América Latina, a nivel mundial el mejor proyecto de M-learning según (Unesco, 2013) es el proyecto Nokia MoMath de Sudáfrica, el mismo que utiliza la función de SMS (servicio de mensajes cortos) de los teléfonos móviles estándar para que los alumnos accedan a contenidos y sea apoyo de la materia de matemáticas (Isaacs, 2012).

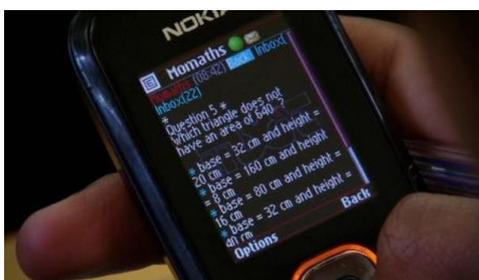


Figura 7 Proyecto MoMath

Fuente: (Carlos, 2010)

En la Figura 7, se observa, que el proyecto MoMath no cuenta con la tecnología móvil más actual, encuentra la manera de usar los celulares básicos. Un estudio realizado a este proyecto señala que cuanto más los alumnos utilizan el servicio y realizan los ejercicios en un 100% son mucho más efectivos en los resultados finales, lo que demuestra que conservan lo que han aprendido y no simplemente memorizan.

Los docentes implicados en este proyecto querían que el proyecto no termine en la fase de prueba debido a los resultados exitosos que obtuvieron con el mismo (Unesco, 2013).

De la recopilación de información de todos los proyectos en América Latina y el más exitoso a nivel mundial, se concluye que el aprendizaje móvil alcanza a personas marginadas (zonas rurales y comunidades indígenas), además, que con el uso de celulares con bajas características tecnológicas se obtuvieron mejores resultados en el aprendizaje del estudiante, aún más se puede obtener con el uso de tecnología moderna, pero siempre resaltando que la tecnología por sí sola no obtiene cambios, depende mucho de la iniciativa en cada proyecto que se emprenda para obtener buenos resultados.

(Unesco, 2013) Menciona la transición hacia el libro de texto digital es una de las tendencias más consolidadas del aprendizaje móvil, en los próximos quince años no cabe duda de que el aprendizaje móvil se integrará más en el sistema educativo ordinario.

Así como en la actualidad se observa que las computadoras son indispensable en la educación, pronto las tecnologías móviles serán de uso indispensable en la educación formal e informal.

1.3 Minería de datos.

Actualmente existen grandes cantidades de datos: impresos y/o digitales, que son obtenidos de una rutina sin tener en cuenta. Si se realizara un análisis correspondiente a estos datos, se pudiese obtener información relevante, aquí surge el término de minería de datos, el cual permite analizar y explorar la información, obteniendo un nuevo conocimiento a partir del mismo.

Se puede creer que al realizar un análisis estadístico se puede llegar a resolver los mismos problemas que con minería de datos pero como lo menciona (Ramón, 2014) la minería de datos es descendiente de la estadística, presentando diferencias entre ellas: (Ramón, 2014)

- La minería de datos permite mejorar en rendimiento, adaptabilidad y optimización de tiempo.
- En cuanto a la generación de un modelo, en estadística se puede validar un modelo ya generado más no crearlo, en cambio con la minería de datos se puede generar el modelo, es así que la estadística usa un método científico y técnicas confirmatorias para demostrar una hipótesis previamente elaborada. En cambio la minería de datos descubre hipótesis de los datos generados por el problema, permitiendo validar la hipótesis al contrarrestarla con otros datos.

(Moine, Haedo, & Gordillo, 2013) Menciona que la minería de datos (MD) es el proceso de extraer conocimiento útil, comprensible y novedoso de grandes volúmenes de datos o también llamado Data Mining.

Según (Hand, Mannila , & Smyth, 2001) la MD, es el análisis de grandes cantidades de datos, que reúne a varias ciencias (Estadísticas, Informática, Matemáticas, Ingeniería, etc.) por consiguiente se puede extraer información de gran valor que se encontraba oculta en toda la recopilación de información.

Una forma más clara para poder comprender lo que significa minería de datos, la menciona (Pérez López, 2007) donde expone que la minería de datos es el núcleo del proceso de descubrimiento de conocimientos, para este proceso de extracción de conocimiento KDD (Knowledge Discovery in Databases) se tiene un esquema como se observa en la Figura 8:

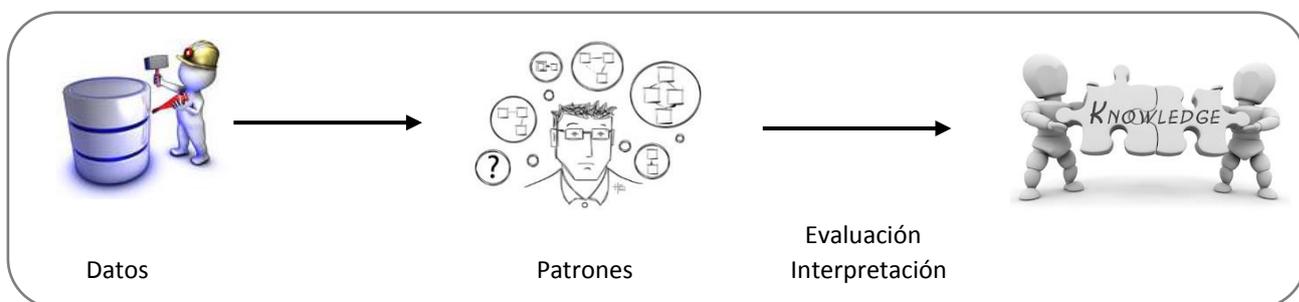


Figura 8 Esquema del Proceso KDD

Fuente: Elaboración propia

No se debe olvidar, lo que menciona (Riquelme, Ruiz , & Gilbe, 2006), que no es lo mismo la minería de datos que el KDD, ya que la minería de datos es una parte del KDD. La MD aplica algoritmos para la obtención de patrones a partir de datos, es decir construye modelos donde se realiza la descripción de los patrones y su relación, en cambio el KDD se centra en el proceso global de descubrir conocimiento de grandes volúmenes de datos, incluyendo el almacenaje, acceso a los datos e interpretación de los patrones obtenidos de la minería de datos.

El KDD consta de las siguientes fases:

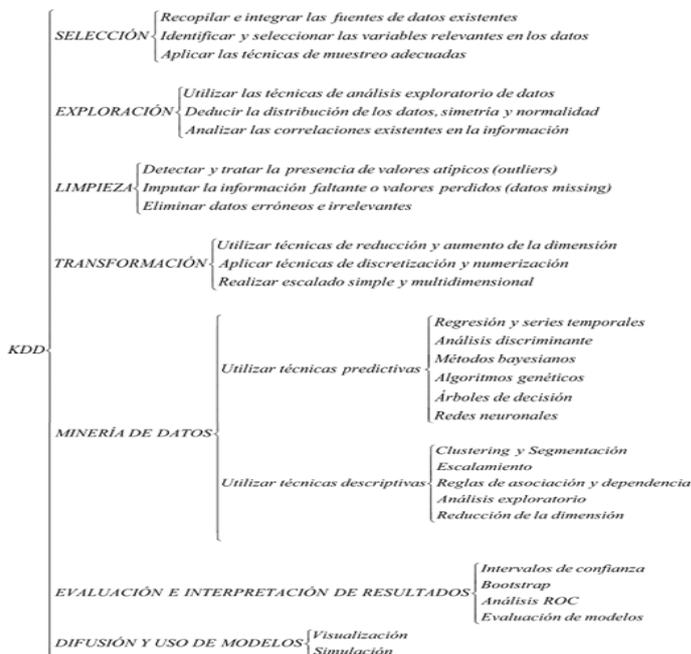


Figura 9 Fases del proceso de extracción del conocimiento

Fuente (Pérez López, 2007)

Es por esto que Pyle (2003 citado en Moine, Haedo, & Gordillo, 2013) plantea que no es recomendable trabajar directamente con los datos porque primero se debe establecer la problemática, las personas que intervienen, las expectativas y las necesidades del usuario, debido a que se debe tener claro el fin y funciones que cumplirá el proyecto.

1.3.1 Tareas Y Métodos.

Tomando en cuenta lo que menciona (Hernández Orallo, Ramírez Quintana, & Ferri Ramírez, 2004; IBM, 2014) la diferencia entre una tarea y un método, es que la tarea integra la información para iniciar el proceso de preparación del modelo de minería de datos. Se distinguen varios tipos de tareas, cada una se la puede considerar como un tipo de problema que será resuelto con un algoritmo, esto significa que cada tarea tiene sus requisitos, mientras que un método o técnica son aquellos que sirven para solucionar una tarea.

A continuación veremos los tipos de tareas que existen según (Pérez López, 2007) & (Hernández Orallo, Ramírez Quintana, & Ferri Ramírez, 2004) :

1.3.1.1. Predictivas o Supervisadas.- son tareas en las que se va a predecir uno o más valores para uno o más ejemplos. Se tiene presente la correspondencia entre los ejemplos y los valores de salida. Se pueden definir varias tareas predictivas:

- Clasificación (o discriminación).
- Clasificación suave.
- Estimación de probabilidad.
- Categorización.

- Preferencias o priorización.
- Regresión.

1.3.1.2. Descriptivas o No Supervisadas.- el objetivo es describir los datos existentes. A continuación se mencionan las tareas descriptivas:

- Agrupamiento (clustering).
- Correlaciones y factorizaciones.
- Reglas de asociación.
- Dependencias funcionales.
- Detección de valores e instancias anómalas.

Cada una de las tareas ya mencionadas necesita de métodos, técnicas o algoritmos para resolverlas, es importante mencionar que una tarea puede ser resuelta con varias técnicas distintas, pero una técnica puede resolver varias tareas.

La clasificación de las técnicas o métodos es:

- Técnicas algebraicas y estadísticas.
- Técnicas bayesianas.
- Técnicas basadas en conteos de frecuencias y tablas de contingencia.
- Técnicas basadas en árboles de decisión y sistemas de aprendizaje de reglas.
- Técnicas relacionales, declarativas y estructurales.
- Técnicas basadas en redes neuronales artificiales.
- Técnicas basadas en núcleo y máquinas de soporte vectorial.
- Técnicas estocásticas y difusas.
- Técnicas basadas en casos, en densidad o distancia.

Con todas las clasificaciones mencionadas tanto de las tareas como métodos se puede observar en la Tabla 4, la correspondencia entre tareas y técnicas:

Tabla 4 Correspondencia entre tareas y técnicas

TÉCNICAS		TAREAS				
		PREDICTIVO			DESCRIPTIVO	
		Clasificación	Regresión	Agrupamiento	Reglas de Asociación	Correlaciones/ Factorizaciones
Basadas en redes neuronales artificiales	Redes Neuronales	✓	✓	✓		
	Redes de Kohonen			✓		
Basadas en árboles de decisión y sistemas de aprendizaje de reglas	Árboles de decisión ID3, C4.5, C5.0	✓				
	CN2 rules (cobertura)	✓			✓	
	Árboles de decisión CART	✓	✓			
	Otros árboles de decisión (J48)	✓	✓	✓	✓	
Algebraicas y estadísticas	Regresión lineal y logarítmica		✓			✓
	Análisis factorial y de comp. Ppales.					✓
	Regresión logística	✓			✓	
	Análisis discriminante multivariante	✓				
Basadas en casos, en densidad o distancia	Kmeans			✓		
	Twostep, Cobweb			✓		
	Algoritmos genéticos y evolutivos	✓	✓	✓	✓	✓
	Vecinos más próximos	✓	✓	✓	✓	
	Basadas en conteos de frecuencias y tablas de contingencia	A priori				✓
Bayesianas	Naive Bayes	✓				
Basadas en núcleo y máquinas de soporte vectorial	Máquinas de vectores soporte	✓	✓	✓		

Fuente: (Hernández Orallo, Ramírez Quintana, & Ferri Ramírez, 2004)

1.3.2 Herramientas.

Las herramientas de minería de datos permiten predecir futuras tendencias y comportamientos, brindando ayuda en la toma de decisiones, basándose en la información

recolectada. Teniendo en cuenta que las herramientas usan algunas técnicas o métodos para obtener modelos o asociaciones.

Las herramientas de minería de datos pueden responder a preguntas de negocios que tradicionalmente consumen demasiado tiempo en ser resueltas y en las cuales los usuarios de la información, no están dispuestos a aceptar. Estas herramientas exploran las bases de datos en busca de patrones ocultos, encontrando información predecible que un experto no puede llegar a encontrar porque se encuentra fuera de sus expectativas (Llor Díaz, 2006).

Existen varias herramientas para realizar minería de datos como lo menciona (García González, 2013):

1.3.2.1 XLMiner.

Es un complemento para Excel, con funcionamiento mediante macros, basado en métodos de aprendizaje de estadística.

Posee 5 grandes grupos de operaciones (Frontlinessolver, 2015):

- Particiones
- Clasificación
- Predicción
- Análisis Affinity
- Series de tiempo
- Reducción de datos y exploración

Según (Frontlinessolver, 2015) existen ventajas y desventajas de esta herramienta.

Ventajas:

- Posee un buen manual que se encuentra dentro de la sección ayuda, que describe los distintos métodos y parámetros.
- Presenta opciones de configuración y trabajo (interface) amigables para cada método.
- Los formatos de presentación de resultados como gráficos tablas e indicadores de cambios son muy ordenados.
- Existe una gran cantidad de videos tutoriales para los distintos métodos en los que indican paso a paso qué se necesita hacer.

Desventajas:

- XLMiner, es accesible como herramienta de prueba. (Periodo limitado de 30 días).
- Limita el tamaño de la base de datos.

- No posee indicadores de errores claros. Como XLMiner trabaja en base a macros de Excel, al parametrizar alguna operación con datos inadecuados, la operación se interrumpe, siendo imposible de recuperar los últimos cambios realizados.

1.3.2.2 IBM SPSS STATISTICS

SPSS Statistics, es una extensa plataforma que está diseñado para llevar la inteligencia a las decisiones de los individuos, los grupos, los sistemas y la empresa. Ofreciendo técnicas de modelado univariado y multivariado, utilizando técnicas analíticas para extraer conocimiento de los datos (IBM, 2016).

Posee un conjunto de datos y herramientas de análisis predictivo y descriptivo teniendo modelos de clúster, regresión general, árboles de decisión además de consumir un conjunto de reglas y modelos de apoyo de máquinas de vectores (López García, 2012).

Finalmente y al igual que en los programas anteriores se trata de un programa comercial y se debe pagar una licencia para su uso.

1.3.2.3 SAS Enterprise Miner.

Modelado descriptivo y predictivo proporciona una visión que impulsa una mejor toma de decisiones. Puede agilizar el proceso de minería de datos para desarrollar modelos rápidamente (Sas, 2014).

La solución Enterprise Miner se basa en la metodología SEMMA (Sample, Explore, Modify, Model, Assess) desarrollada por SAS Institute y puede definirse de la siguiente forma:

- Muestra (Sample)
- Explora (Explore)
- Modificar (Modify)
- Modelo (Model)
- Evaluar (Asses)

Para diseños de modelado predictivo, el rendimiento de cada modelo y los supuestos del modelo pueden ser verificados a partir de las parcelas de predicción y gráficos de diagnóstico; las relaciones clave entre éstos permiten encontrar los patrones más importantes.

Según (Sas, 2014) las principales características de esta herramienta son:

- GUI fácil de usar y el procesamiento por lotes.
- Preparación de datos sofisticados, el resumen y la exploración.
- El modelo predictivo y descriptivo.

- La integración de código abierto con R.
- Prestaciones de alto rendimiento.
- Una manera rápida, fácil y autosuficiente para usuarios de negocios para generar modelos.
- Comparaciones de modelos, de informes y de gestión.
- Puntuación automatizada.
- Procesamiento escalable.

Tomando en cuenta todas sus características concluimos que es una herramienta de gran potencia al momento de trabajar con grandes bases de datos, sin embargo el alto precio de su licencia hace difícil su obtención.

1.3.2.4 Salford Systems Data Mining.

Es una empresa que se dedica al desarrollo de software de minería de datos ofreciendo los siguientes productos (Salford Systems, 2015):

- Software CART: posee una variedad de análisis de alta calidad en la minería de datos, es la única basada en árboles de decisión ofreciendo una clasificación multi-plataforma robusta.
- TreeNet: es una herramienta de minería de datos más flexible y potente de Salford, capaz de generar consistentemente modelos extremadamente precisos. Ofrece modelos predictivos de alto rendimiento. El sistema de detección de interacción no sólo ayuda a mejorar el rendimiento del modelo, sino que también ayuda en el descubrimiento de valiosos nuevos segmentos y patrones no reconocidos previamente.
- MARS: cuando se prefiere resultados en una forma similar a la regresión tradicional es ideal esta herramienta, el enfoque MARS a un modelo de regresión descubre difíciles patrones y relaciones para otros métodos de regresión, por lo tanto rastrea cualquier patrón detectado en datos, sirve como herramienta de exploración inicial de los datos.
- RandomForests: esta herramienta aprovecha la potencia de alternativas múltiples de análisis, sus puntos fuertes los grupos de proximidad, predicción de los resultados futuros, la identificación de predictores importantes, descubrir patrones de datos, sustituyendo los valores perdidos con imputaciones, y con gráficos interesantes.

La suma de las predicciones realizadas a partir de árboles de decisión determina la predicción general, incorporando nuevos análisis de clúster de métrica libre.

1.3.2.5 Tariykdd.

Puede definirse como una herramienta genérica de tareas múltiples débilmente acoplada a un SGBD. TariykDD comprende cuatro módulos que cubren el módulo de conexión, tanto a

archivos planos como a bases de datos relacionales, un módulo de utilidades con clases y librerías comunes a toda la aplicación, un módulo kernel que incluye las tareas de pre procesamiento, algoritmos de asociación, clasificación y rutinas de visualización de resultados. Este es una herramienta de software totalmente libre.

1.3.2.6 WEKA.

Waikato Environment for Knowledge Analysis, es una herramienta que contiene una biblioteca de librerías JAVA para la extracción de conocimiento desde bases de datos.

Es un software desarrollado en la Universidad de Waikato (Nueva Zelanda) bajo licencia GNU-GPL lo cual ha impulsado que sea una de las herramientas más utilizadas en el área en los últimos años. Se trata de una herramienta de gran potencia (Weka, 2013).

Tiene las siguientes ventajas:

- Interfaz gráfica muy sencilla pudiendo aplicar, evaluar con un gran número de algoritmos de minería a grandes conjuntos de datos.
- Constituida por una serie de paquetes de código abierto con diferentes técnicas de pre procesado, clasificación, agrupamiento, asociación y visualización.
- Posee un formato propio que le permite crear de forma rápida los dataset necesarios.
- Permite aplicar filtros ya sea a los atributos o a las instancias.
- Debido a su variedad de algoritmos implementados permite realizar tareas con sus respectivos características de datos como: clustering, clasificación, asociación, regresión y visualización.

En la siguiente tabla se observa la comparación de varias herramientas a fin de determinar cuál es la mejor para poder obtener los patrones de comportamiento, se analizan 4 herramientas de pago y 2 herramientas gratuitas donde se evalúa varias de las siguientes características funcionales como tipo de licencia, la implementación de algoritmos, interfaz, plataforma, visualización de datos, tipos de modelos y métodos, entre otras.

Tabla 5 Comparación de Herramientas

Características	XLMiner	IBM SPSS	SAS Enterprise Miner	Salford Systems Data Mining	Tariykdd	Weka
Licencia Libre	no	no	no	no	si	si
Requiere conocimientos avanzados	no	no	no	no	no	no
Accesos a SQL	si	si	si	si	si	si
Multiplataforma	si	si	si	si	si	si
Requiere base de datos especializados	-----	no	-----	si	no	no

Métodos de máquinas de soporte vectorial	si	si	si	si	no	si
Métodos Bayesianos	si	si	si	si	no	si
Puede combinar modelos	si	si	si	-----	no	si(no resulta muy eficiente)
Modelos de clasificación	si	si	si	si	si	si
Implementación de árboles de decisión	si	si	si	si	si	si
Modelos de regresión	si	si	si	si	no	si
Clustering y agrupamiento	si	si	si	no	no	si
Interfaz amigable	si	si	si	-----	si	si
Permite visualización de datos	si	si	si	si	si	si

Fuente: Adaptado de (Vallejo & Tenelanda, 2012)

Luego de haber comparado características funcionales se observa que las herramientas de pago en comparación con las de licencia libre poseen iguales características con la diferencia que las herramientas de pago es más factible realizar la combinación de modelos es así que entre todas las herramientas las que más cumple con los parámetros planteados son: en cuanto a licencia libre WEKA y de pago SPSS.

1.4 Metodología para proyecto de minería de datos.

Existen algunas metodologías para proyectos de minería de datos como lo menciona Moine, Haedo y Gordillo (2013) en un estudio comparativo de las metodologías para proyectos de minería de datos, comparando Semma, Crisp - Dm y Catalyst. Donde los autores concluye que solo Crisp – Dm y Catalyst son considerados una metodología de minería de datos, debido a que cuentan con la descripción de las actividades específicas de cada fase brindando una guía paso a paso para llevar acabo el desarrollo de la metodología.

1.4.1 Comparación de metodologías

Para poder elegir una de las mejores metodologías se tomará en cuenta el estudio de Moine, Haedo y Gordillo (2013) donde evalúan dos de las metodologías: Crisp – Dm y Catalyst realizando una comparación de cada una de la actividades específicas que componen las fases generales. En la siguiente tabla se visualiza un resumen de dicha comparación:

Tabla 6 Comparación de Metodologías

Fase	Crisp-Dm	Catalyst
Análisis del problema	6/9 (66%)	7/9 (77%)
Selección y preparación de los datos	3/5 (60%)	4/5 (80%)
Modelado	3/4 (75%)	3/4 (75%)

Evaluación	4/4 (100%)	4/4 (100%)
Implementación	4/4 (100%)	3/4 (75%)
Total	20/26 (77%)	21/26 (80%)

Fuente: (Moine, Haedo, & Gordillo, 2013)

El resultado de la tabla indica que ambas metodologías cumple con la mayoría de actividades específicas (descritas en el Anexo), siendo ambas una buena opción al momento de ser aplicadas dependiendo del contexto del problema.

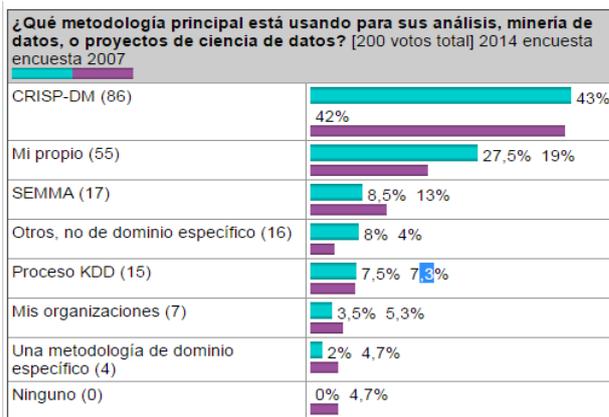


Figura 10 Encuesta Metodología de Minería de Datos

Fuente: (Kdnuggets, 2014)

1.4.2 Crisp – Dm (Chapman, y otros, 2000).

Crisp – Dm (Cross Industry Standard Process for Data Mining) es una guía de referencia creada en 1999, recordando los resultados de la figura 1:10 donde se puede ver que en comparación con otras metodologías ésta se convirtió en la más utilizada para el desarrollo de proyectos de minería de datos tomando en cuenta que su éxito se debe a que está basada en experiencias reales.

La metodología Crisp – Dm se encuentra descrita en un modelo de proceso jerárquico, en un conjunto de tareas de cuatro niveles yendo de lo general a lo específico como se observa en la siguiente imagen:

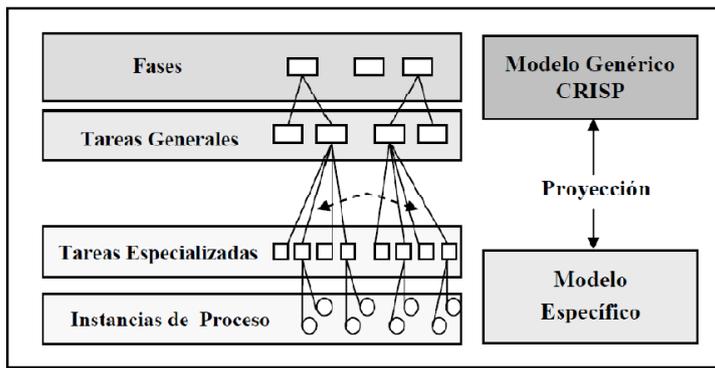


Figura 11 Modelo Crisp - DM
Fuente: (Chapman, y otros, 2000)

La diferencia entre el modelo genérico y específico es que el modelo genérico consta de 2 niveles, en el primer nivel se encuentran 6 fases (aunque en el gráfico original figuran 3 fases), en el segundo nivel se encuentran tareas genéricas o generales que corresponde a las 6 fases del primer nivel. En el modelo específico tenemos el tercer nivel el mismo que describe cómo las acciones de tareas generales deben ser realizadas en situaciones específicas llamadas tareas especializadas. Finalmente en el cuarto nivel se trata de un registro de acciones, decisiones y resultados de una minería de datos real las cuales se denominan instancias del proceso.

Se debe tener en cuenta que el modelo no es lo mismo que el ciclo de vida, el modelo brinda una perspectiva general del ciclo de vida en cambio en el ciclo de vida se detallan las fases de Crisp - Dm además de las relaciones entre ellas.

Por lo tanto el ciclo de vida que presenta Crisp – Dm posee fases, con sus respectivas tareas y la relación correspondiente.

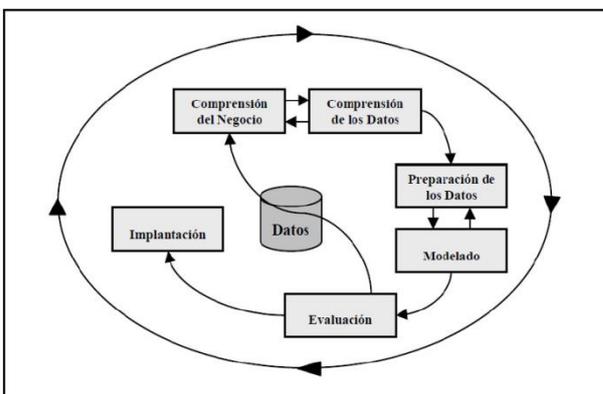


Figura 12 Ciclo de Vida Crisp - DM
Fuente: (Chapman, y otros, 2000)

Crisp –Dm posee seis fases y cada una de ellas posee sus respectivas tareas, como se detallan a continuación:

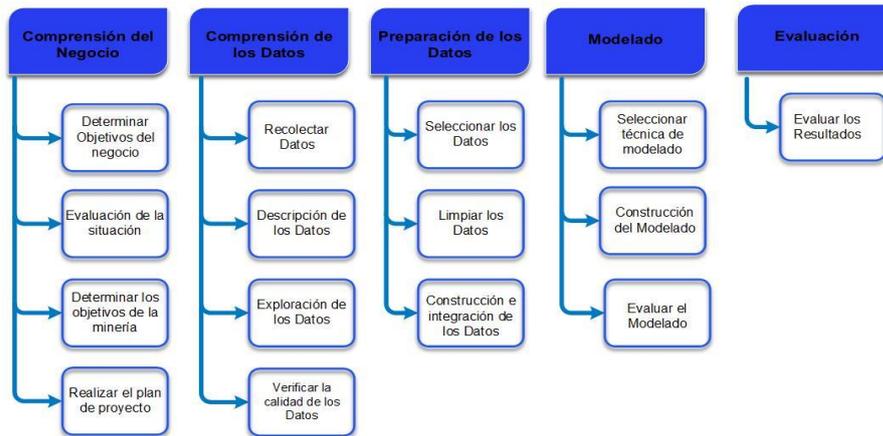


Figura 13 Fases Crisp - DM
Adaptado: (Chapman, y otros, 2000)

1.4.2.1. Descripción de fases Crisp – DM (Chapman, y otros, 2000).

1.4.2.1.1. Comprensión del negocio

Probablemente es una de las fases con mayor relevancia porque agrupa tareas de comprensión de objetivos y de requisitos del proyecto, permitiendo recolectar datos correctos y así poder realizar una correcta interpretación de resultados.

Las respectivas tareas que se deben desarrollar junto con los entregables de cada una de ellas las vemos a continuación:

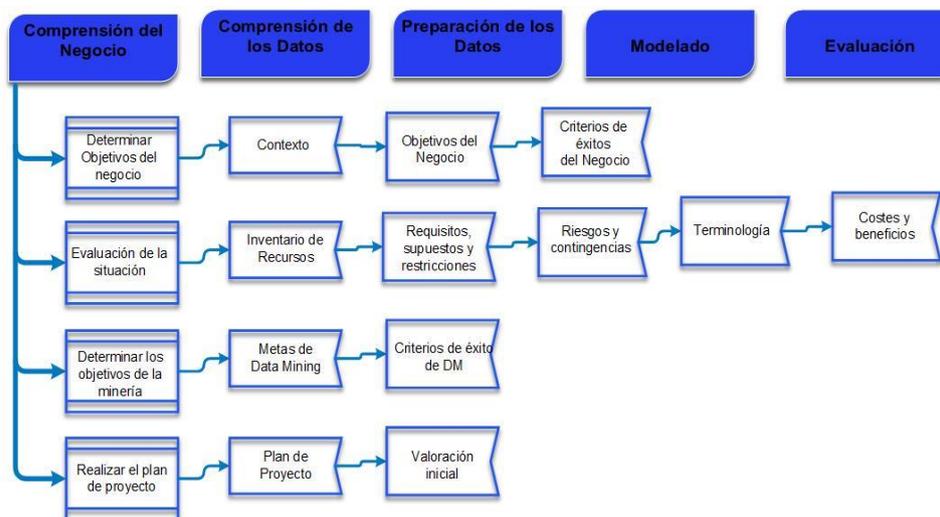


Figura 14 Comprensión del Negocio
Adaptado: (Chapman, y otros, 2000)

1.4.2.1.2. Comprensión de los Datos

Esta fase realiza la recolección inicial de datos para tener un primer contacto con el problema y familiarizarse con los mismos, determinando la calidad y estableciendo relaciones evidentes para plantear las primeras hipótesis.

Posee las siguientes tareas junto con los entregables de cada una de ellas:

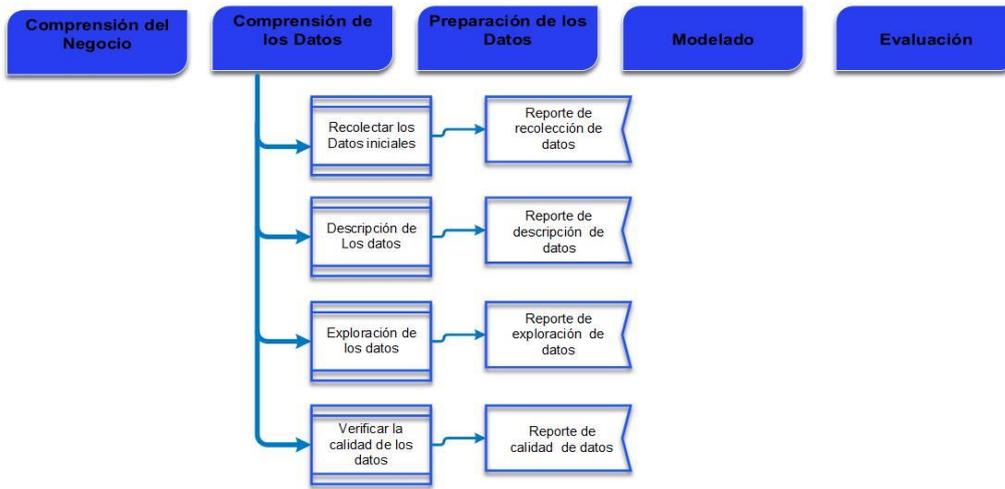


Figura 15 Comprensión de Datos

Fuente: (Chapman, y otros, 2000)

1.4.2.1.3. Preparación de los Datos

Esta fase prepara los datos para la adaptación de técnicas de Minería de Datos que se empleen luego, incluye tareas generales de selección de datos para lo que se aplicará una técnica de modelado, limpieza de datos, determinación de variables adicionales y cambios de formato de datos.

Tomando en cuenta que con la fase de modelado se encuentran muy relacionada debido a que según la técnica de modelado elegida los datos necesitaran ser procesados de diferente forma.

A continuación podemos observar las tareas junto con los entregables de cada una de ellas:

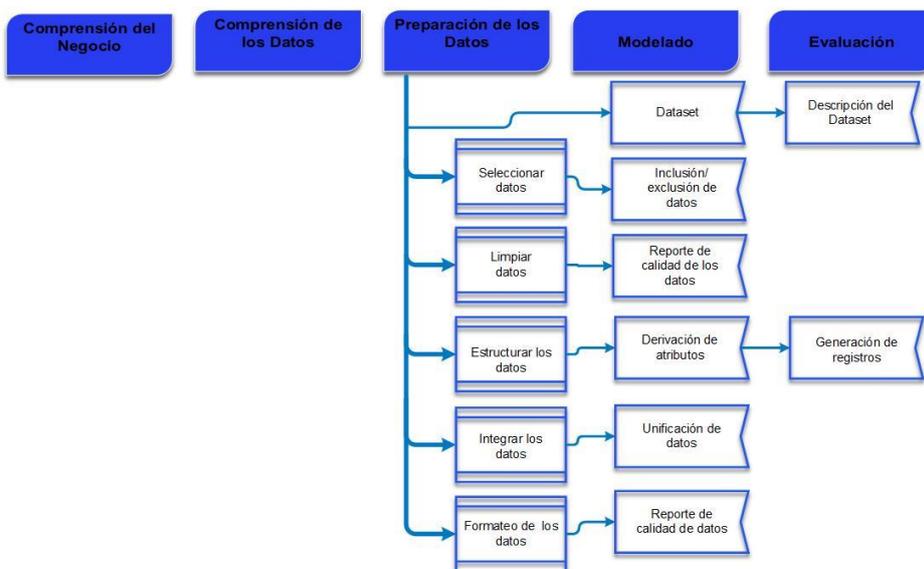


Figura 16 Preparación de Datos

Fuente: (Chapman, y otros, 2000)

1.4.1.1.4. Modelado

Se deben elegir técnicas de modelado apropiadas para el proyecto tomando en cuenta los siguientes criterios según (Chapman, y otros, 2000):

- ✓ Ser apropiada al problema.
- ✓ Disponer de datos adecuados.
- ✓ Cumplir los requisitos del problema.
- ✓ Tiempo adecuado para obtener un modelo.
- ✓ Conocimiento de la técnica.

Antes de la generación y evaluación del modelo se necesita determinar un método de evaluación del modelo, teniendo en cuenta que el modelo dependerá de las características de los datos y de la precisión que se desea obtener con el modelo.

Para poder hacer una correcta selección de la técnica del modelado se considera el objetivo principal del proyecto y la relación con las herramientas de Minería de Datos.

En la siguiente imagen observamos las tareas junto con los entregables de cada una de ellas:

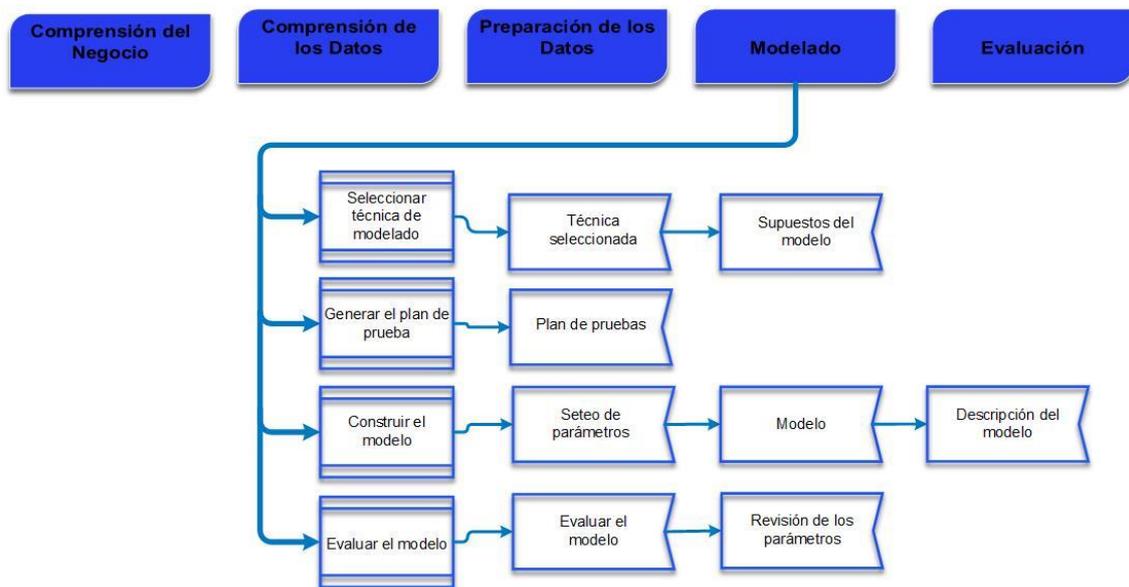


Figura 17 Modelado

Fuente: (Chapman, y otros, 2000)

1.4.1.1.4. Evaluación

Existen dos opciones para realizar la evaluación del modelo la primera es evaluar el porcentaje de respuesta del modelo a los objetivos de negocio, y la segunda es probar los modelos sobre aplicaciones de prueba en la aplicación real.

También se debe tomar en cuenta evaluar el modelo con respecto a objetivos diferentes a los originales brindando la posibilidad de que surja información adicional.

Las tareas junto con los entregables de cada una de ellas son las siguientes:

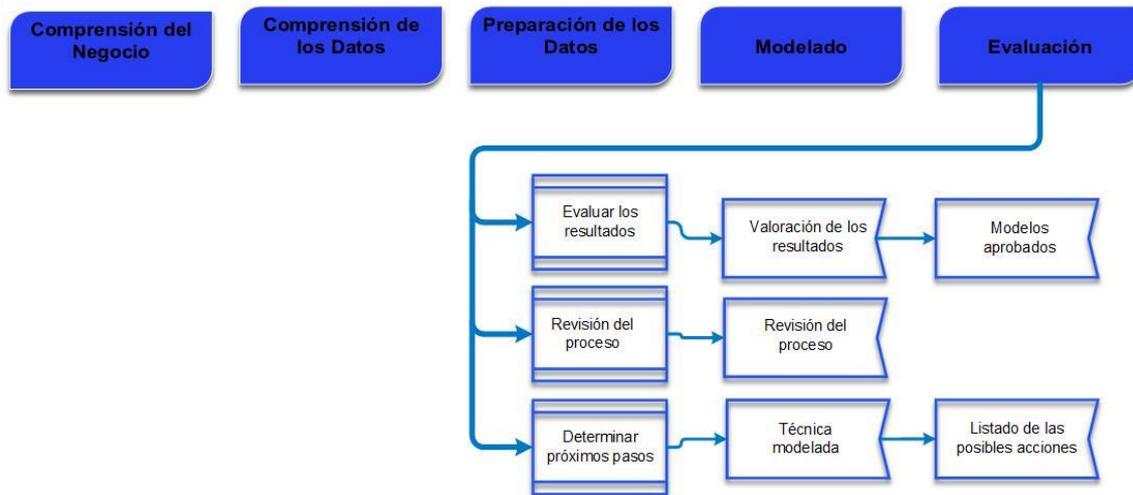


Figura 18 Evaluación
Fuente: (Chapman, y otros, 2000)

1.5 Proyectos relacionados

La minería de datos es un técnica muy usada en la actualidad, a continuación se destacan los proyectos más actuales dentro la literatura revisada en cuanto a la implementación de m-learning en un contexto universitario tomando en cuenta las estrategias de recolección de datos como la minería de datos o técnicas estadística.

1.5.1 Patrones de M – Learning en el aula virtual. (López Hernández & Silva Pérez, 2014)

Contexto: Se usaron datos de la interactividad de los estudiantes de Pregrado en la plataforma de la universidad además de encuesta a 460 estudiantes de la modalidad presencial para la obtención de los patrones M – Learning.

Descripción: El proyecto busca evaluar la importancia que tienen los dispositivos móviles para el aprendizaje en la educación superior e identificar los principales patrones de uso, para lo cual usa dos instrumentos de recolección de datos: minería web del uso en la plataforma virtual de la universidad mediante el cual exploraron las tendencias del uso de esta nueva tecnología en los últimos cuatro cursos académicos además de identificar los principales patrones de comportamiento con la ayuda de una encuesta.

La técnica que aplicaron fue la de agrupación y clasificación para poder distribuir similar de los estudiantes en grupos homogéneos.

Los resultados que obtuvieron fueron los siguientes:

Minería Web.- obtuvieron que el 25% de las entradas al sistema LMS (Learning Management Systems) se realizan con dispositivo móvil en la ocasión que más ingresan es en periodo de exámenes superando el 30% de entradas al sistema además del acceso levemente mayor en horas nocturnas con dispositivos móviles.

Encuesta.- el resultado de la encuesta es el 75% de los estudiantes utilizan dispositivos móviles con fines de aprendizaje, donde solo un 7.3% posee algún dispositivo móvil, además de la predisposición de los estudiantes para usar dispositivos móviles en su aprendizaje a pesar de que las universidades no facilitan el acceso a los mismos.

1.5.2 Uso de Ebooks por los alumnos de la universidad de Andrews: un estudio de las actitudes, percepciones y comportamientos (Marques de Oliveira, 2012)

Contexto: Se usaron 960 estudiantes de la Universidad de Andrews en Estados Unidos de la modalidad presencial, para la obtención de actitudes, percepciones y comportamientos de los mismos.

Descripción.- El problema a resolver con este estudio es la identificación de patrones de uso y actitudes de los estudiantes hacia los Ebooks, para lo cual se aplica la técnica de una encuesta a los alumnos para observar la preferencia entre los libros de texto impresos o los libros digitales (Ebooks).

Esta investigación se realiza con el fin de entender mejor como los estudiantes de la Universidad de Andrews utilizan los Ebooks. La encuesta se la realizó enfocada en el conocimiento, uso, razones por las que elegir o no los Ebooks, opinión sobre lo que haría a los Ebooks más adecuado en su área de estudio y cuáles son los factores que ellos piensan que dificultan su uso.

Los resultados presentados son dos grupos de usuarios:

- Los estudiantes que adquirieron un libro digital
- Estudiantes prefirieron el formato de impresión tradicional.

En cuanto al uso y frecuencia, los resultados mostraron que un 40% de los estudiantes habían utilizado anteriormente los Ebooks manifestando su experiencia como positiva, afirmando que se trata de fácil navegación y menos libros que llevar. Sin embargo el restante 60% de los estudiantes reportaron haber tenido una experiencia negativa, de los cuales el 26.8% indicaron que el libro impreso es mejor para leer, el 11.3% que era difícil desplazarse, existiendo otras razones como dificultades al subrayar, tomar notas dificultando el aprendizaje.

Las razones para no usar Ebooks según los estudiantes encuestados, La gran mayoría de los estudiantes (87,9%) encuestados no eligen la versión digital del libro ¿Cuándo se preguntó por qué, más de la mitad de estos estudiantes (59,2%) declararon que preferían la versión impresa, ya que no estaban familiarizados sobre cómo usarlo, además de opinar que los libros electrónicos son demasiado caros, así como las dificultades de no saber cómo encontrar libros electrónicos aplicables a su investigación.

1.5.3 El uso del libro electrónico en una biblioteca universitaria para docentes, alumnos de pregrado y postgrado: actitudes y comportamientos de los usuarios (Allen Shelburne, 2013)

Contexto: Se tomó en cuenta docentes y alumnos de pregrado y postgrado de la modalidad presencial de la universidad de Illinois, aplicando una encuesta para identificación de patrones.

Descripción.- En este estudio buscaron actitudes y comportamientos de los estudiantes que utilizan libros electrónicos en una biblioteca de la universidad de Illinois, donde se observa lo que al estudiante le gusta o disgusta de los libros electrónicos.

La técnica aplicada fue una encuesta a gran escala diseñada para investigar los patrones y actitudes de los estudiantes.

La encuesta revela que los estudiantes poseen la predisposición para adaptar Ebooks en su educación por lo que se han vuelto una servicio importante en la biblioteca, pero se resalta algo importante en la investigación la falta de variedad de Ebooks en la biblioteca además de tomar en cuenta que para cuando requieran leer un Ebook completo prefieren combinar tanto el libro físico como el electrónico.

Los estudiantes comentan varias ventajas encontradas al momento de utilizar Ebooks como las que vemos a continuación:

- Cuida el medio ambiente.
- Leer en cualquier lugar (portabilidad).
- Búsqueda rápida en el libro.
- Subrayar y escribir notas de lo relevante del texto.
- Pre – formato de referencias de los libros.

Además de obtener datos estadísticos donde el docente usa el 91% un Ebook para investigación, el 83% de estudiantes graduados utilizan para la investigación; y el 69% de los usuarios de pregrado los utilizan para la investigación con estos datos reiteramos como el uso de Ebooks está creciendo y las personas se adaptan perfectamente a su uso.

CAPÍTULO II: ANÁLISIS DE LA PROBLEMÁTICA Y DISEÑO DE LA SOLUCIÓN

2.1 Análisis de la problemática

En la actualidad, la tecnología en general es el eje principal para el desarrollo de nuestras actividades cotidianas, aún más la tecnología móvil volviéndose indispensable como se lo menciono en el capítulo I según la revista International Journal of Mobile y Blended Learning. Es así como en la educación, la tecnología móvil está cambiando parte del modelo educativo con una de sus innovaciones como los Ebooks (libros electrónicos), los cuales brindan al estudiante una manera interactiva en el proceso de aprendizaje pero tomando en cuenta que para generar ayuda en el aprendizaje del estudiante debe cumplir un ecosistema M – Learning como se lo menciono en el capítulo I.

Muchas Universidades están implantando esta tecnología, como bibliotecas digitales que en si es un catálogo el cual permite acceder a la información de la biblioteca desde cualquier lugar o en celulares implementando aplicaciones que sirven de refuerzo para el estudiante.

La diferencia que marca el proyecto EBOOKS de la Universidad Técnica Particular De Loja es que su objetivo principal es la entrega de todo el material bibliográfico en formato digital a perpetuidad, así como la oportunidad de acceder a un dispositivo de lectura, brindándoles la ventaja de contar con su material en todo momento, facilitando la lectura y la oportunidad de aprovechar al máximo su tiempo disponible para realizar su tareas.

Como se mencionó en el capítulo I se debe tomar en cuenta el ecosistema que plantea (Woodill, 2011). En la UTPL la ejecución del proyecto ebooks, ha generado el inicio de un ecosistema M – Learning, sin embargo hay elementos que faltan por desarrollar (líneas entrecortadas) como se muestra en el gráfico siguiente:

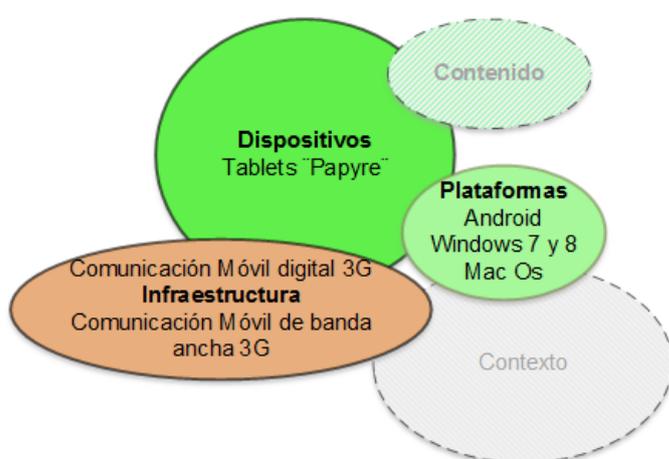


Figura 19 Ecosistema M –Learning UTPL

Fuente: Elaboración Propia

Como se observa aún el ecosistema de Ebooks – UTPL no es completo debido a la falta de contenidos y contextos innovadores que generen un conocimiento adicional del interés del estudiante.

El desarrollo de estos contenidos y contextos innovadores debe estar basado en patrones de comportamiento de los estudiantes en un contexto M – Learning, a fin de generar contenidos adaptados al perfil y necesidades de cada estudiante.

Si bien existen estudios relacionados, éstos son enfocados a educación primaria y secundaria y a contextos de consulta en bibliotecas o en el aula presencial, aspectos muy diferentes a nuestra realidad, que es un contexto de educación superior, con modalidad de estudios a distancia y entrega general gratuita a estudiantes de primer ciclo.

Uno de los inconvenientes para el presente estudio es que el registro de interacción de los estudiantes en cuanto al uso de las Tableta digital se la tiene desde el presente ciclo Octubre – Febrero 2016 además no a todos se les entregó a tiempo su dispositivo por ende no se tiene de todos los estudiantes de primer ciclo el registro de interacción.

Ante este problema se determinó obtener patrones de comportamiento de los estudiantes que son parte del proyecto Ebooks, mediante la Tableta digital comparando con el comportamiento que tienen en el EVA, a través de técnicas de minería de datos

También se realizó una encuesta la cual permitió obtener información adicional que no se registra en la Tableta digital ni en el EVA.

2.2 Diseño de la solución

Para el desarrollo de la solución se aplicará minería de datos para obtener un modelo descriptivo, que permita saber cómo interactúa el estudiante con la Tableta digital y el EVA, el mismo que se encuentra conformado por patrones de comportamiento de los estudiantes de la Modalidad Abierta y a Distancia de la UTPL.

El modelo usará los resultados de las variables de interacción de la Tableta digital y el EVA además de las respuestas de la encuesta obteniendo un modelo descriptivo correcto que ayudará a saber las necesidades de los estudiantes de la Modalidad Abierta y a Distancia de la UTPL.

Pero para saber que metodología es la correcta se tomó en cuenta el análisis comparativo de Moine, Haedo y Gordillo (2013) que se lo indico en el capítulo I.

Donde si bien los resultados del análisis comparativo la metodología Catalyst obtuvo un 80% y Crisp – Dm obtuvo un 77% en relación al cumplimiento de sus actividades específicas de las fases donde el 3% menos que obtuvo Crisp – Dm corresponde a las fases de:

Análisis del problema donde Crisp – Dm no propone una evaluación de las fuentes de datos en esta fase, pero lo hace luego de realizar el plan del proyecto, también se toma en cuenta que no se analizan todas las posibles soluciones al problema. Pero en cambio Catalyst no define un criterio de éxito para el proyecto en el cual se prevé los posibles problemas y sus respectivas soluciones, ni tampoco realiza una evaluación general de las técnicas de minería que se podrían utilizar.

Por lo tanto en esta fase de análisis del problema la mejor metodología sería Crisp – Dm debido a su definición de criterios de éxito además de la evaluación de las técnicas de minería de datos.

Selección y preparación de los datos Crisp – Dm no propone actividades para la revisión del conjunto de datos final con el usuario como lo menciona Moine, Haedo y Gordillo (2013), pero esta actividad no es necesaria para la obtención de nuestro modelo debido a que en la planificación de criterios de éxito se toma en cuenta los datos que se necesita para obtener un proyecto exitoso.

Modelado es esta fase Crisp - Dm planifica la forma en la que se evaluarán los resultados finales y Catalyst no lo hace.

Finalmente se tomó la decisión de utilizar la metodología Crisp – Dm debido a la definición de criterios de éxito, evaluación de técnicas de minería, evaluación de resultados finales y también considerando resultados de la encuesta realizada por (Kdnuggets, 2014) a sus usuarios donde Crisp – Dm es la metodología mayor usada con resultados exitosos.

A continuación se resume el diseño de la solución donde necesitaremos determinar las variables descriptivas luego, se agruparán los datos según su variable.

Para realizar minería de datos se usará la herramienta WEKA y SPSS donde se ejecutará la técnica de clustering obteniendo el modelo descriptivo el cual nos permitirá la obtención de los patrones de comportamiento de los estudiantes, se utilizarán ambas herramientas para determinar con cuál de ellas se pueda relacionar el comportamiento que tienen en el EVA con el de la Tableta digital.

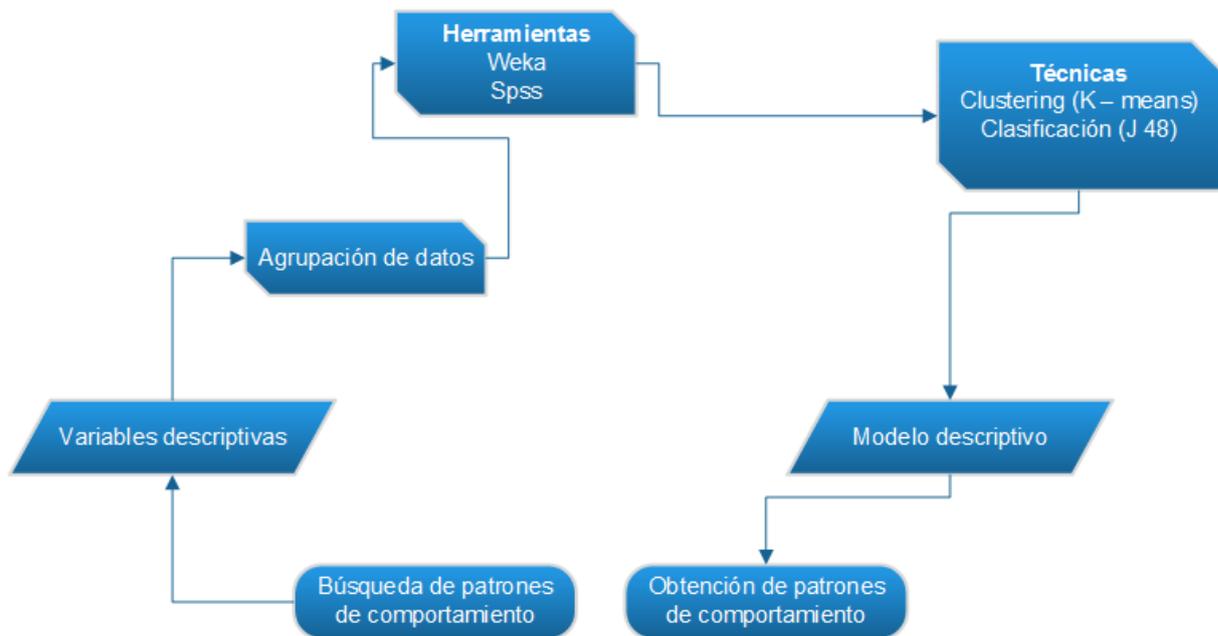


Figura 20 Solución
Fuente: Elaboración Propia

2.2.1 Variables

Luego del análisis de la problemática se han determinado variables vinculadas con los ámbitos: datos personales, EVA, Tableta digital. A continuación se presentan la distribución de las variables:

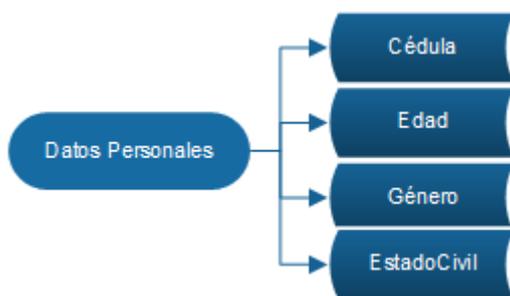


Figura 21 Variables de Datos Personales Modelo Descriptivo
Fuente: Elaboración Propia

En Cuanto a las variables del EVA se recolectaron todas las interacciones de las actividades de los estudiantes en las diferentes herramientas como: cuestionarios (quiz), tareas (assign), foro (forum), chat (chat), recursos (resource), blog (blog).

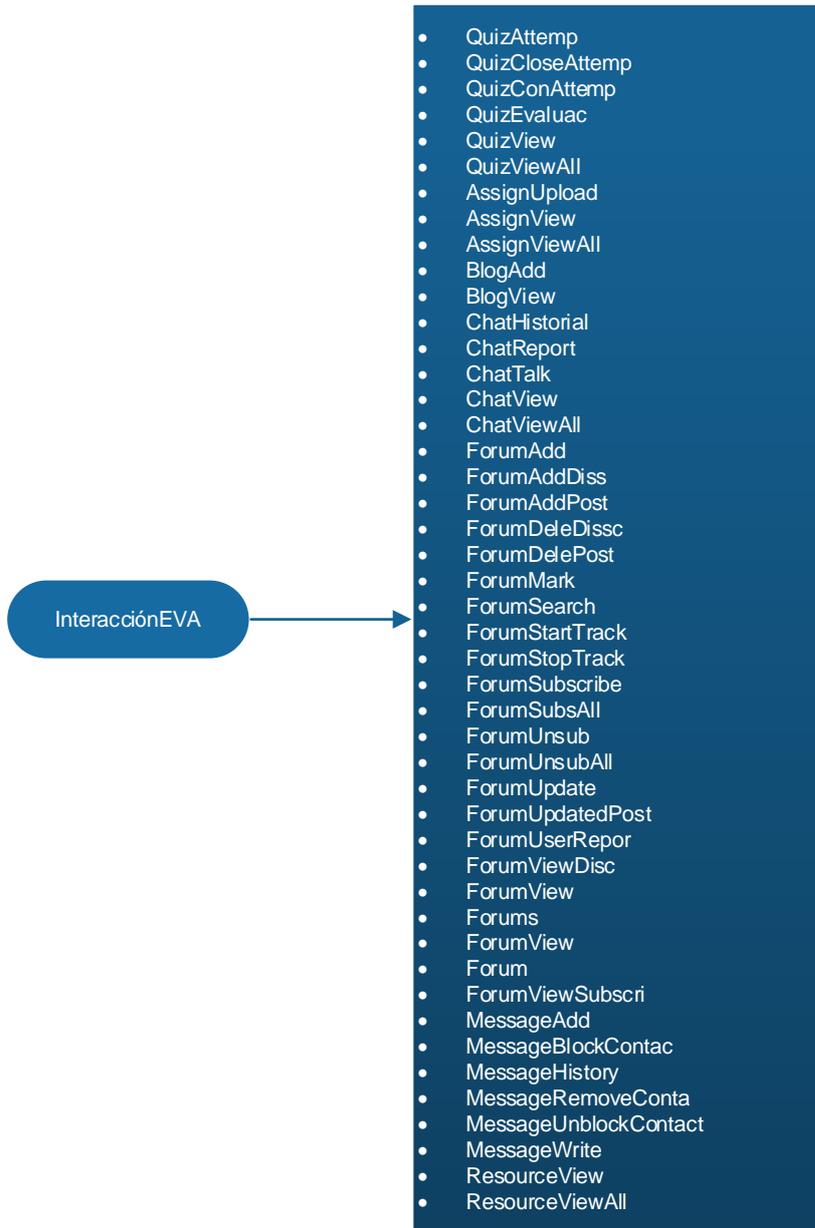


Figura 22 Variables de Interacción en el EVA

Fuente: Elaboración Propia

En cuanto a la interacción en las Tableta digital la data se la clasificó en 3 categorías (aplicaciones, juegos y descargas) el criterio de clasificación se lo detalla en el capítulo 3.

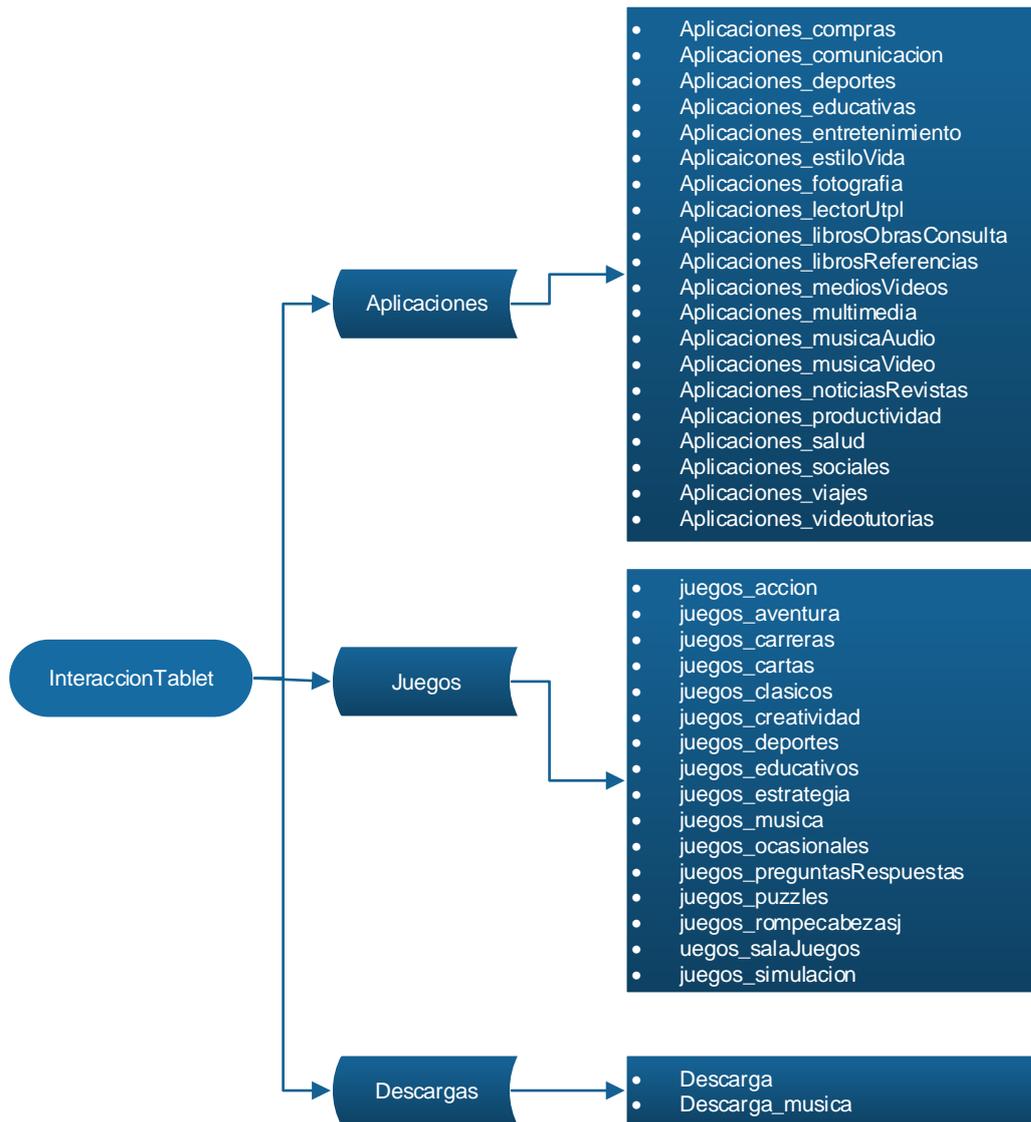


Figura 23 Variables de Interacción en las Tableta digital
 Fuente: Elaboración Propia

2.2.2 Herramienta de minería de datos

Como se mencionó en el Capítulo I, al momento de realizar las comparaciones entre las herramientas que existen para minería de datos se concluyó que existen dos buenas herramientas como lo son WEKA y SPSS, se realizarán pruebas en ambas para determinar en cuál de ellas, además de obtener el comportamiento individual en la Tableta digital y en el Eva, se pueden comparar estos comportamiento y saber qué relación tiene el uno con el otro.

2.2.3 Técnicas de minería de datos

Para saber que técnicas implementar, en la sección de diseño de la solución, se tomó en cuenta los siguientes criterios:

- No necesitar el ingreso de alguna condición para la definición de patrones .

- Agrupación de técnicas específicas para modelos descriptivos ver tabla 4.
- Agrupamiento según características de los estudiantes y posibles variables que influyan en un determinado comportamiento.

Luego de haber realizado la tabla de tareas y técnicas de minería de datos se determinó el uso de las técnicas de agrupamiento (clustering), la mismas que tienen como objetivo la agrupación de datos que formen clusters o conglomerados que sean homogéneos internamente y heterogéneos externamente, además de tener la característica de un aprendizaje no supervisado esto permitirá la definición de los patrones de comportamiento de los estudiantes.

Clustering K – means

Según (Perversi, 2007) es un algoritmo inductivo no supervisado que utiliza el método de agrupamiento y su objetivo es encontrar grupos de observación con características semejantes, no posee una variable dependiente.

Este algoritmo distingue los prototipos o centros, buscando agrupar aquellos prototipos que contengan características semejantes, donde el objetivo es tratar de disminuir la varianza total o función de error cuadrático para que el algoritmo pueda generar los mejores resultados.

$$V = \sum_{i=0}^k \sum_{j \in S_i} |x_j - \mu_i|^2$$

Figura 24 Error Cuadrático

Fuente: (Perversi, 2007)

K-means comienza particionando los datos en k subconjuntos no vacíos, aleatoriamente o usando alguna heurística, esto significa que usarán criterios, reglas o principios para obtener los mejores grupos de clustering . Luego calcula el centroide de cada partición como el punto medio del clúster y asigna cada dato al clúster cuyo centroide sea el más próximo. Luego los centroides son recalculados para los grupos nuevos y el algoritmo se repite mientras sigan habiendo la aparición de patrones y cuando no haya más datos que cambien de grupo de una iteración a otra (Perversi, 2007).

Dado k (número de clúster), el algoritmo K-means sigue el siguiente proceso Ale (como se citó en Rüegg, Gutiérrez, 2008):

1. Particionar los objetos en k subconjuntos no vacíos.
2. Computar los centroides de los clusters de la partición corriente. El centroide es el centro (punto medio) del clúster.
3. Asignar cada objeto al clúster cuyo centroide sea más cercano.

4. Volver al paso 2, parar cuando no haya más reasignaciones.

Se tiene que evaluar que el resultado del error cuadrático en los clúster tomando en cuenta que entre menor sea el error cuadrático cada clúster será mas similar.

Algoritmo J48

Sierra (2006 citado en Wilford Rivera, Rosete Suárez, & Rodríguez Díaz, 2008) menciona que es una implementación del algoritmo C4.5.

J48 empieza seleccionando el atributo del cual empezará a dividir la muestra de entrenamiento más conocida como nodo raíz, con el propósito que en los sub conjuntos exista una mínima diferencia respecto a la clase o más conocidas como hojas.

El proceso antes mencionado es recursivo es decir una vez que encontró la variable que cumple el mayor grado de homogeneidad con respecto a la clase, se repite el proceso para cada uno de los nodos hijos. Lo que determinará que se finalice el proceso es el criterio del método de pre-poda y post-poda evitando que las condiciones sean muy específicas y den un mal comportamiento al modelo tomando en cuenta que el modelo es una aproximación de lo que se busca.

El proceso de post-poda lo realiza por medio del método de reducción del error (reducedErrorPruning) o las divisiones sobre variables discretas sean binarias (binarySplits).

(Wilford Rivera, Rosete Suárez, & Rodríguez Díaz, 2008) indica las siguientes propiedades:

- Admite atributos nominales y numéricos.
- Se permiten ejemplos con valores desconocidos.
- El criterio de división está basado en la entropía y la ganancia de información.

El árbol que se genera se conforma de:

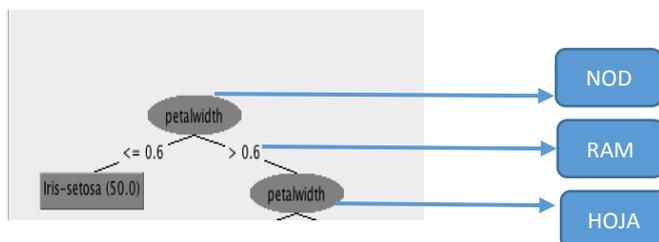


Figura 25 Partes de un Árbol J48
Fuente: Elaboración propia

CAPÍTULO III: PROCESO DE MINERÍA DE DATOS

Introducción

Se utilizará la metodología Crisp Dm luego de haber realizado una comparación de metodologías en el capítulo 1.4 donde la que más se adapta al problema es Crisp Dm, la misma que permitirá obtener un modelo descriptivo del comportamiento de los estudiantes de la titulación de Derecho del periodo Octubre – Febrero 2016.

3.1 Fase I. Compresión del negocio

3.1.1 Determinación de los Objetivos del Negocio

3.1.1.1 Contexto

Se debe tomar en cuenta que la población pertenece a la modalidad a distancias de primer ciclo de la titulación de derecho los cuales todos poseen una Tableta digital Digital correspondiente.

3.1.1.2 Objetivos del negocio

La Universidad Técnica Particular de Loja entrega a los estudiantes de Modalidad Abierta y a Distancia el material bibliográfico de forma digital mediante Ebooks el mismo que posee los siguientes objetivos:

- Dar aprendizaje M – learning.
- Ayudar al medio ambiente.
- Mejorar la alfabetización digital.
- Brindar contenido que mejore el aprendizaje y se adapte al estudiante.

3.1.1.3 Criterios de éxito del negocio

Obtención de un modelo descriptivo del comportamiento de los estudiantes.

Desarrollar el caso de estudio mediante base de datos del EVA, Tableta digital y encuesta aplicada.

Interpretar los resultados que se obtengan en WEKA y SPSS para obtener los patrones de comportamiento de los estudiantes.

3.1.2 Evaluación de la Situación

Luego de haber realizado el análisis de la problemática se ha podido observar la situación actual para la ejecución del proyecto permitiendo plantear los requerimientos, presunciones y restricciones que se observan a continuación:

3.1.2.1 Requerimientos

Para el desarrollo del proyecto se requiere lo siguiente:

- Acceso a la información del Eva y las Tableta digital.

- Manejo de Sql, WEKA Y SPSS.
- Respuesta de los estudiantes a encuesta planteada.
- Evaluación de los resultados por parte de los expertos del negocio.

3.1.2.2 Presunciones

- A todos los estudiantes del periodo octubre – febrero 2016 se les entrego la Tableta digital.
- La evaluación a distancia a través del EVA la deben realizar obligatoriamente todos los estudiantes.
- Los datos de los estudiantes deben ser correspondientes al periodo octubre – febrero 2016.
- Los estudiantes poseen un perfil relacionado entre el EVA y la Tableta digital.
- En cuanto a las aplicaciones que brinda la universidad en las Tableta digital, deberían ser usadas por todos los estudiantes.

3.1.2.3 Restricciones

- En los datos demográficos recolectados de los estudiantes matriculados en primer ciclo del periodo octubre – febrero 2016 no se cuenta con los datos de ingresos económicos ni el área en la que trabajan, los cuales se los habría podido utilizar para determinar si intervienen o no en el comportamiento del estudiante.
- No se puede acceder a la interacción dentro de cada una de las aplicaciones dentro de la Tableta digital debido a que no son propiedad de la universidad.
- Se puede tomar la información solo de aquellos estudiantes que registran la recepción de la Tableta digital, esto debido a que si se toman en cuenta los datos de los estudiantes que si tiene interacción en el EVA pero no han recibido la Tableta digital no habrá datos que relacionar.

3.1.2.4 Recursos

Los recursos que tienen disponibles son:

- **PERSONAL**

Tabla 7 Recursos Personales

Recursos Personal

Expertos del negocio.

Administrador del Eva.

Administrador del sistema de gestión de Ebooks.

Expertos de minería de datos.

Fuente: Elaboración propia

- **INFORMACIÓN**

Tabla 8 Recursos de Información

Recursos - Información

- Listado de estudiantes del proyecto Ebooks matriculados en primer ciclo del periodo octubre - febrero 2016, que ya hayan recibido la Tableta digital.
- Información personal de los estudiantes del proyecto Ebooks octubre – febrero 2016.
- Información de las actividades en el EVA de los estudiantes de la carrera de mayor afluencia de listado de estudiantes del proyecto Ebooks octubre – febrero 2016.
- Información del uso de aplicaciones en la Tableta digital de los estudiantes de la carrera de mayor afluencia de listado de estudiantes del proyecto Ebooks octubre – febrero 2016.

Fuente: Elaboración propia

- **SOFTWARE**

Tabla 9 Software a Utilizar

Herramienta	Uso
XAMPP	Levantamiento del gestor de base de datos
Postgress	Herramienta para el manejo, limpieza y consulta de la base de datos del uso de las aplicaciones en la Tableta digital.
HeidiSQL	Herramienta para el manejo, limpieza y consulta de la base de datos de las actividades realizadas en el EVA.
Excel	Herramienta para el tratamiento de datos.
WEKA	Herramienta para técnicas de minería de datos.
SPSS Statics	Herramienta para análisis estadístico y para la aplicación de técnicas de minería de datos

Fuente: Elaboración Propia

3.1.2.5 Terminología

3.1.7.1 Terminología del negocio

- ✓ EVA.- Entorno virtual de aprendizaje.

- ✓ Proyecto Ebooks.- Proyecto en el que la universidad entrega Tableta digital a los estudiantes con una primera titulación en la universidad con sus respectivos libros digitales.
- ✓ UTPL.- Universidad Técnica Particular de Loja.
- ✓ Papyre Destokp.- Aplicación que posee varios servicios que brinda la universidad para el estudiante.
- ✓ M-Learning.- Aprendizaje a través de tecnología móvil.
- ✓ Titulación.- Hace referencia a las distintas carreras que la universidad oferta.

3.1.7.2 Terminología de la minería de datos

- ✓ Iteración.- Número de veces que se repite un proceso para encontrar lo deseado.
- ✓ Dato nominal.- No siguen un orden determinado y se pueden realizar igualdades y desigualdades.
- ✓ Dato ordinal.- Poseen un orden definido y se pueden realizar comparaciones de mayor y menor que, al mismo tiempo igualdades y desigualdades.
- ✓ Dato Atípico.- Es un dato muy distante del resto de datos.

3.1.8 Determinación de los Objetivos de la Minería

- ✓ Encontrar patrones de comportamiento de los estudiantes que son parte del proyecto Ebooks.
- ✓ Identificar grupos de estudiantes según su interacción en el EVA como en la Tableta digital.
- ✓ Identificar si existen variables que influyen en su comportamiento.

3.1.3 Plan del Proyecto

Se estimó 9 meses para la ejecución del proyecto, la misma que se encuentra por componentes o fases teniendo en cada una un entregable como se lo observa en la siguiente tabla:

Tabla 10 Plan del Proyecto

Componente	Tiempo	Entregables
Estado del Arte: m-learning, proceso de minería de datos, técnicas de minería de datos.	1 mes	Informe con el estado del arte
Comprensión del problema: análisis de la necesidad, situación o contexto y objetivos	2 mes	Informe con: Los criterios de éxito del problema. Requisitos, supuestos y restricciones

de la minería de datos		Criterios de éxito de la minería de datos. Plan del proyecto
Comprensión de los datos: recolección, descripción, exploración y verificación de los datos	1 mes	Reporte de recolección, descripción, exploración y calidad de los datos.
Preparación de los datos: Selección, limpieza estructuración, integración y formateo de los datos	2 meses	Informe donde se presente la: recogida, extracción, limpieza de los datos a analizar
Modelado: Selección de la técnica de modelado, diseño del plan de pruebas, construcción y evaluación del modelo.	2 meses	Informe con la descripción del modelo y resultado de la evaluación
Evaluación de los resultados:	1 meses	Informe con los modelos aprobados y posibles acciones a implementar
TOTAL	9 meses	

Fuente: Elaboración Propia

3.2 Fase II. Comprensión de los datos

En esta fase de deber tomar en cuenta los siguientes aspectos para decidir qué tipo de muestreo se va a utilizar:

- Se pretende encontrar patrones de comportamiento de los estudiantes que utilizan la Tableta digital entregada por la universidad.
- El registro de los datos se la tiene desde el periodo octubre - febrero 2016.
- No todos los estudiantes recibieron a tiempo su dispositivo.
- No todas las carreras poseen un número alto en cuanto a la entrega de la Tableta digital.

Se puede observar que aunque el proyecto Ebooks tiene algún tiempo en la universidad el registro de datos se la posee solo del último ciclo.

3.2.1 Muestreo

Para saber qué tipo de muestreo se va a utilizar es necesario tomar en cuenta las características que menciona (Celestino, Flores & Rangel, 2004):

Muestreo Probabilístico

Su principal característica es que cada elemento de la población tiene la misma probabilidad de ser elegida para la muestra.

Muestreo no Probabilístico

En cambio este muestreo no asegura que cierta parte de la población sea excluida por ende no se puede generalizar los resultados para toda la población.

Se puede concluir que debido a los aspectos antes mencionados en cuanto a restricciones de toda la población del proyecto Ebooks, se ve necesario usar el muestreo no probabilístico porque no se puede generalizar que las características sean las mismas en toda la población debido a que no se tiene acceso a toda la población sin embargo del conjunto de elementos que se posee se debe tomar una muestra representativa como lo indica (López, 2004).

Se encuentran ya planteados criterios que se deben tomar en cuenta al tomar la muestra, en consecuencia se usará el muestreo por criterios que menciona (López, 2004) el mismo que se caracteriza por escoger su muestra basándose en algunos criterios establecidos como es el caso de la presente investigación.

Es por ello que se analizó el listado de estudiantes pertenecientes al Proyecto Ebooks siendo un total de 8034 distribuidos de la siguiente manera:

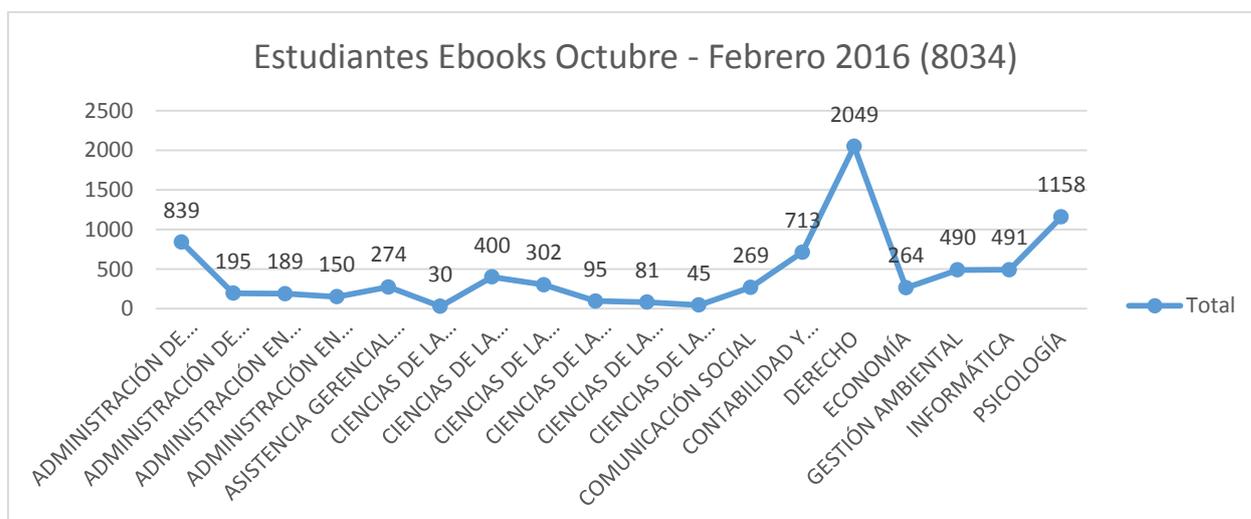


Figura 26 Afluencia de estudiantes del proyecto Ebooks
Fuente: Elaboración Propia

Ahora se evaluará de estos 8034 cuantas Tableta digital han sido entregadas a tiempo antes de iniciar el periodo académico como se lo observa en el siguiente gráfico:

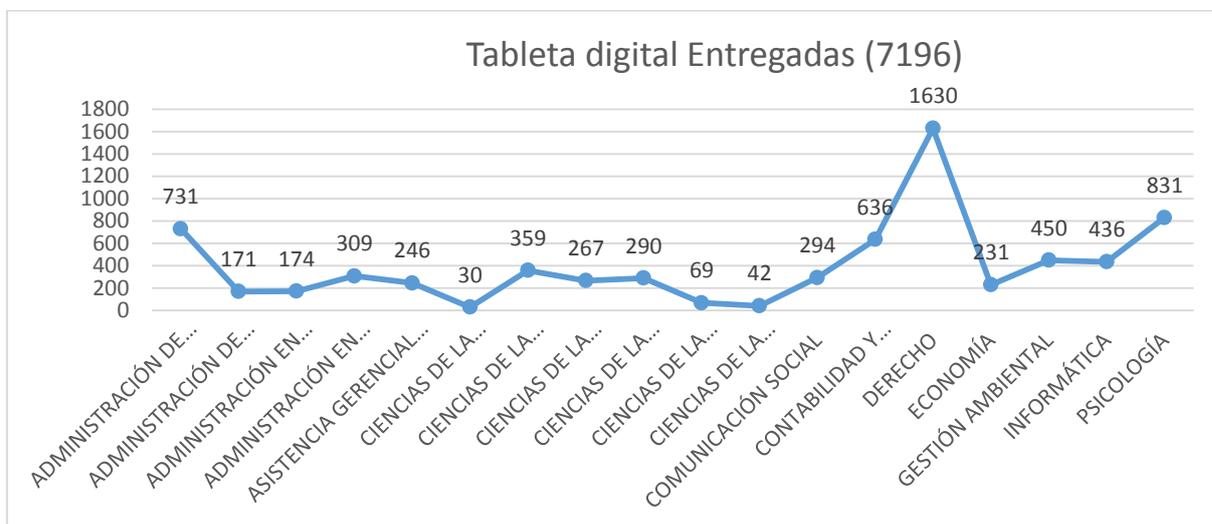


Figura 27 Número de Tablet Entregadas a los Estudiantes
Fuente: Elaboración Propia

Se observa en el figura 27 que han sido entregadas 7196 Tableta digital a los estudiantes del proyecto Ebooks estos 7196 estudiantes serán nuestra población.

Ahora para determinar el tamaño de la muestra es necesario tener en cuenta que se trabaja con una población finita es decir que se conoce el tamaño exacto de la población por ende se usará la siguiente fórmula:

$$n = \frac{N \cdot Z^2 \cdot p \cdot (1-p)}{(N-1) \cdot e^2 + Z^2 \cdot p \cdot (1-p)}$$

Figura 28 Fórmula para la obtención del tamaño de la muestral
Fuente:(Ochoa, 2013)

A continuación se desarrollará la fórmula antes mencionada:

$$n = \frac{7196(2.575)^2(0.5)(1 - 0.5)}{(7196 - 1)(0.03)^2 + (2.575)^2(0.5)(1 - 0.5)}$$

$$n = \frac{7196(6.63)(0.5)(0.5)}{(7195)(0.0009) + (6.630625)(0.5)(0.5)}$$

$$n = \frac{11928.49}{6.4755 + 1.65765625}$$

$$n = \frac{11928.49}{8.13}$$

$$n = \frac{11928.49}{8.13}$$

$$n = 1468$$

Como resultado en el tamaño de la muestra de obtuvo 1468, por ende se analiza la población representada en la figura 27 en donde se encuentra representada toda la población determinando que se escogerá a los estudiantes de derecho, debido a que es notable la titulación con mayor cantidad de estudiantes y la que más se acerca al resultado de la muestra obtenida con un total de 1630 estudiantes.

De los cuales se obtendrá la interacción en el EVA como en la Tableta digital del primer bimestre académico.

3.2.1 Recolección de Datos

Para la recolección de datos de usará la información que se genera en la interacción de los siguientes sistemas:

Entorno virtual de aprendizaje (EVA).- Brinda varias opciones a los estudiantes en cada una de las asignaturas como son: participación en foros, chat, resolución de evaluaciones, vista de archivos, etc. Esta base de datos se la almacenó localmente y así mismo para realizar las consultas se lo hizo mediante la herramienta Heidi.

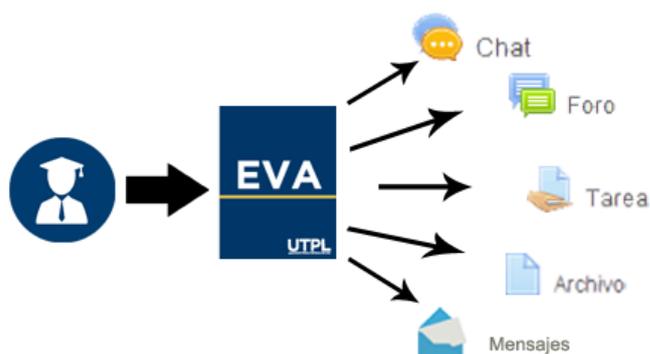


Figura 29 Interacción en el Entorno Virtual (EVA)

Fuente: Elaboración Propia

Sistema Académico (Syllabus).- Mediante este sistema se obtuvieron los datos demográficos aquí no fue necesario acceder a la base de datos ya que fueron entregados en un archivo Excel.



Figura 30 Variables de Datos Personales

Fuente: Elaboración Propia

Servidor Gestión Ebooks.- Para obtener la interacción de cada uno de los estudiantes se accedió por medio de Potsgres Sql a las vistas de información del uso de la Tableta digital.



Figura 31 Interacción en el Servidor de Gestión Ebooks

Fuente: Elaboración Propia

3.2.2 Descripción de los datos

Los datos que fueron proporcionados se encontraron almacenados de la siguiente forma:

- Hoja de excel
- Tablas y vistas relacionadas

A continuación se describen las tablas proporcionadas por cada uno de los sistemas:

Entorno Virtual de Aprendizaje (EVA)

mdl_user.- Posee la información de estudiantes y docentes, la tabla se la puede observar en el [Anexo 1-A].

mdl_log.- Posee información de toda la interacción en el EVA la tabla inicial se encuentra en el [Anexo 1-B].

Finalmente se puede observar en el [Anexo 1-C] la relación entre las dos tablas del EVA.

Sistema Académico (Syllabus)

Los datos iniciales fueron entregados en un archivo excel, la tabla se la puede observar en el [Anexo 2-A].

Servidor Gestión Ebooks (Grammata)

Para acceder a los datos de interacción de las Tableta digital se lo hizo por medio de vistas generadas en el servidor de Grammata, las mismas que se las puede observar en el [Anexo 3-A].

A continuación se detalla los atributos o variables dentro de cada conjunto de datos, tomando en cuenta que se eligieron aquellos atributos con información de actividades comunes de interacción definidas por la institución y que tienen la mayor cantidad de datos:

Entorno Virtual (EVA)

En la tabla de mdl_user no se vio necesario usar las siguientes variables:

- id
- firstname
- lastname
- username
- password
- email
- phone
- address
- city
- firstaccess
- lastaccess

No serán tomadas en cuenta debido a que el id es solamente el identificador de la tabla más no del estudiante, username en cambio es solo el nick del usuario el mismo que no nos sirve para obtener algún patrón.

En cuanto el email, phone, address, city igualmente son variables que no permiten obtener algún patrón de comportamiento y cuya información es ausente en esta tabla. Es por esto que se decidió usar las variables: identificador del usuario y número de cédula para poder relacionarlos con la tabla mdl_log.

Tabla 11 Variables a utilizar mdl_user

mdl_user

userid	Identificador del estudiante en las tablas.
idnumber	Número de cédula del estudiante

Fuente: Elaboración Propia

En la tabla mdl_log no se usarán las siguientes variables:

- id
- ip
- course
- info

Los campos mencionados no serán usados debido a que el id es solo el identificador de la tabla, el campo ip muestra la dirección ip de donde se conecta el estudiante y no es un valor del cual podamos obtener algún patrón de comportamiento, en cuanto a las variables course – info tampoco serán usadas porque únicamente dan la información del identificador de la materia y el estudio no se lo realizará en cuanto a materias sino más bien en cuanto a las acciones que tienen dentro del entorno, la variable info brinda información adicional de esta tabla pero mayormente se encuentra vacía.

A continuación observamos en la tabla 12 las variables a utilizarse:

Tabla 12 Variables a utilizar mdl_log

mdl_log	
userid	Identificador del usuario
module	Actividad
action	Acción

Fuente: Elaboración Propia

Se seleccionó estas variables ya que brindan la información de interacción del estudiante en EVA. Es importante tomar en cuenta que dentro de la tabla mdl_log se tiene diversidad tanto de actividades como acciones determinadas las cuales se las puede observar en el [Anexo 1-D].

Sistema Académico (Syllabus)

En cuanto al archivo excel del sistema académico no se usarán las siguientes variables:

- apellidos
- nombres
- lugar
- modalidad

- mac_fecha_registro
- mac_numero
- titulación
- periodo
- correo_personal
- teléfonos
- dir_domicilio
- colegio
- estado_matricula
- área
- nacionalidad
- provincia
- canton
- etnia
- nivel

No se usarán las variables apellidos, nombres debido a que se encuentran en la tabla del EVA estos datos.

Con respecto a las variables de modalidad, titulación, periodo, estado_matricula, nivel, área, nacionalidad no son necesarias porque se conoce previamente que la muestra pertenece a la modalidad abierta y a distancia de la titulación de derecho de primer ciclo todos con su matrícula legalizada del periodo octubre – febrero 2016 pertenecientes al área Socio-Humanística con nacionalidad Ecuatoriana.

A lo que se refiere a la dir_domicilio, colegio, cantón, provincia, etnia son variables que se encuentran en su mayoría vacías es por ello que no se las puede usar.

Finalmente las variables mac_fecha_registro, mac_numero, correo_personal, teléfonos son variables que no influyen en la generación de patrones de comportamiento.

Tabla 13 Datos personales en la tabla Syllabus

syllabus	
identificación	Cédula del estudiante
fecha_nacimiento	Fecha de nacimiento
genero	Clasificación del género
estado_civil	Clasificación del estado civil

centro Lugar universitario al que pertenece el estudiante

Fuente: Elaboración Propia

En cuanto a las variables a usar del archivo excel del sistema Syllabus en género y estado_civil se tiene los siguientes valores:

- Género:
Femenino
Masculino
- Estado_civil:
Soltero
Casado
Divorciado
Union Libre
Viudo

Servidor Gestión Ebooks (Grammata)

Luego de haber observado las vistas en el [Anexo 3 - A] se determinó que variables se usarán y cuáles no, como se observa a continuación:

biusuario

- nom_usuario
- nombre
- apellidos

Estas variables no se las utilizará porque en el estudio de patrones no se utiliza los nombres propios de cada estudiante, debido a que esto no nos generará ningún patrón.

Las variables que se decidió usar ya que permiten la relación con las otras vistas son las que se muestran a continuación:

Tabla 14 Vista biusuario

biusuario	
id_usuario	Identificador del estudiante en la vista.
cedula	Número de identificación propia del estudiante relacionado con el id_usuario.

Fuente: Elaboración propia

bidispositivo

- num_serie

Esta variable no se la usará debido a que es el número serial de cada dispositivo número de cual no podemos obtener ningún patrón. La variable a utilizar **id_dis_fisico** permite la relación con la vista bidispositivousuario.

Tabla 15 Vista bidispositivo

bidispositivo	
id_dis_fisico	Número de identificación del dispositivo.

Fuente: Elaboración Propia

De las vistas bihistoricoaplicacion y bidispositivousuario se usarán sus dos únicas variables, ya que realizan las vinculaciones entre el dispositivo con el usuario y la aplicación con el dispositivo como se observa en las siguientes tablas:

Tabla 16 Vista bidispositivousuario

bidispositivousuario	
id_dis_fisico	Número de identificación del dispositivo.
id_usuario	Identificador del usuario

Fuente: Elaboración propia

Tabla 17 Vista bihistoricoaplicacion

bihistoricoaplicacion	
id_aplicacion	Identificador de la aplicación.
id_dis_fisico	Identificador del dispositivo.

Fuente: Elaboración propia

En la última vista biaplicacion no se usará la variable sistema_operativo porque la muestra es solo del sistema Android, por ende no es necesario. Las variables que se tomarán en cuenta son el identificador del dispositivo para poder relacionar a que estudiante pertenece el nombre de la aplicación puede observar en la siguiente tabla:

Tabla 18 Vista biaplicacion

biaplicacion	
id	Número de identificación del dispositivo.
nombre	Nombre de las aplicaciones.
paquete	Nombre del paquete de la aplicación

Fuente: Elaboración propia

3.2.3 Exploración de los datos

Una vez seleccionado las variables que se usarán, se obtendrá su distribución. En esta sección se enfocará en las variables edad, género y estado civil, la distribución del resto de variables se presentará en la fase IV porque en esta fase aún no se encuentran categorizadas.

Distribución por edad

La mayor parte de estudiantes tienen 18 años pero se encuentra la particularidad de que existen edades menores a 16 años lo cual no es un valor correcto y serán tratados en el proceso de limpieza de datos.

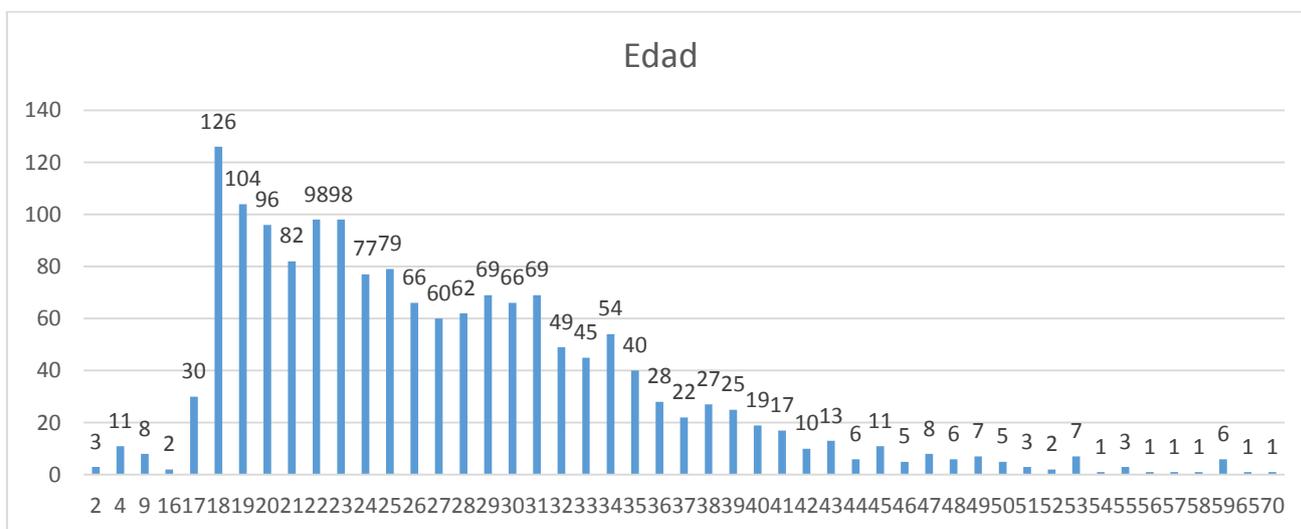


Figura 32 Gráfica de distribución por edades

Fuente: Elaboración propia

Se observa en el gráfico que existen variedad de edades y para poder realizar una mejor distribución es necesario realizar la siguiente categorización:

Tabla 19 Categorización edades

campo	categoría	significado
edad	16_21	16 a 21
	22_27	22 a 27
	28_33	28 a 33
	34_39	34 a 39
	40_45	40 a 45
	46_51	45 a 50
	52_57	52 a 57
	58_63	58 a 63
	64_69	64 a 69
	70_75	70 a 75

Fuente: Elaboración propia

Distribución por estado civil

Se puede observar que la mayoría de estudiantes son solteros con un 64%.

Tabla 20 Distribución estado civil

Estado Civil	Número de Estudiantes
Casado	422
Divorciado	81
Soltero	1027
UnionLibre	63
Viudo	7
Vacio	26
Otro	4
Total general	1630

Fuente: Elaboración propia

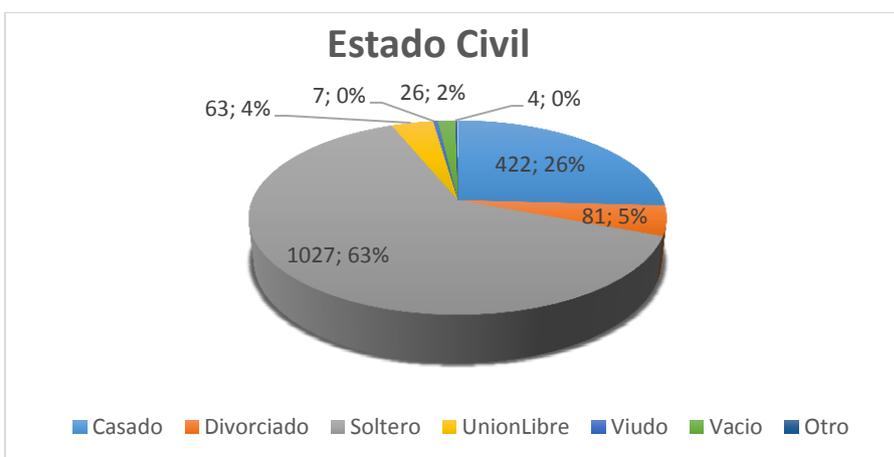


Figura 33 Gráfica de distribución del estado civil

Fuente: Elaboración propia

Distribución por género

En cuanto al género la distribución es mayormente masculino con un 54%.

Tabla 21 Distribución género

Género	Número de estudiantes
Femenino	752
Masculino	878
Total general	1630

Fuente: Elaboración propia

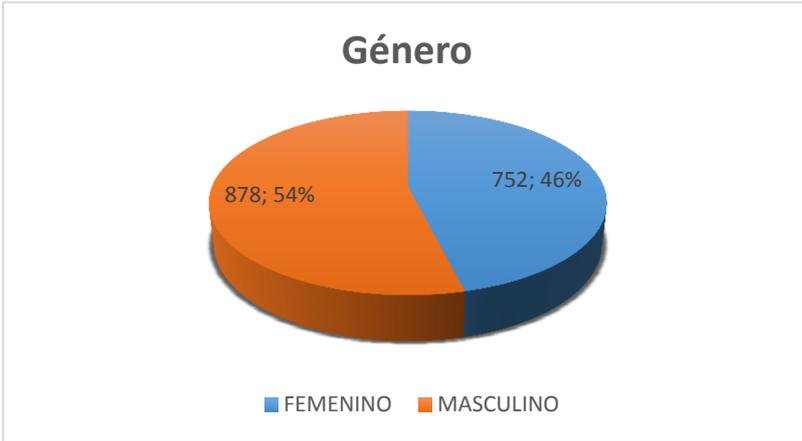


Figura 34 Gráfica de distribución por género
 Fuente: Elaboración propia

Distribución por centro

Se observa que el centro con mayor número de estudiantes es el centro regional Quito con 179 estudiantes.

Debido a la amplitud de la tabla de distribución se la observará en el [Anexo 5], a continuación se observa el gráfico de distribución.

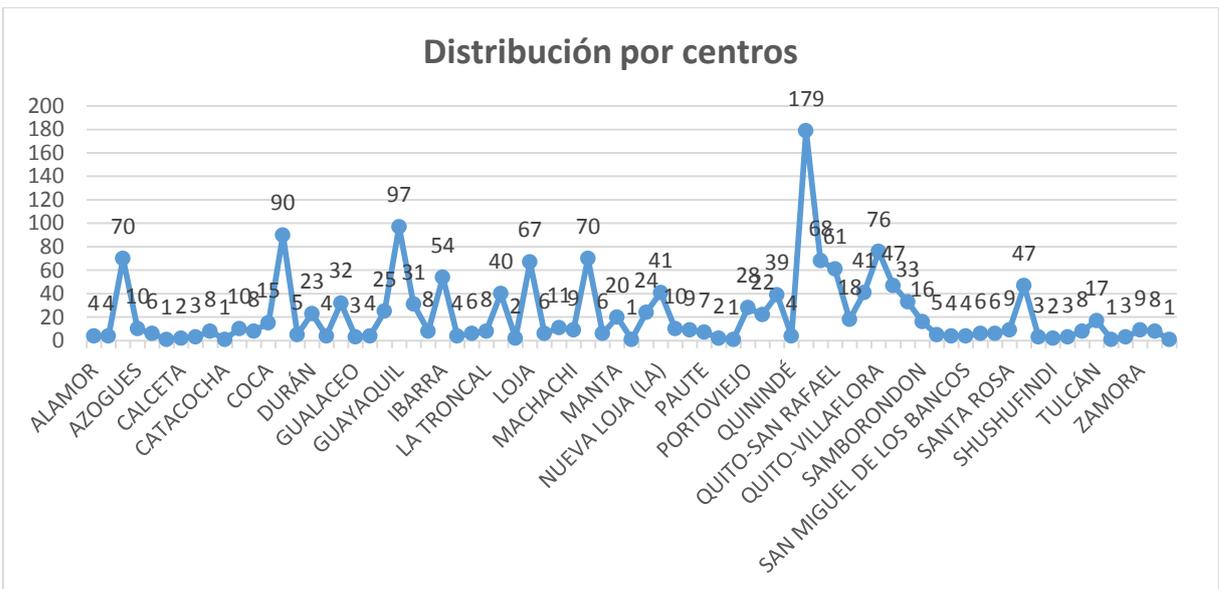


Figura 35 Distribución por centros
 Fuente: Elaboración propia

La distribución del resto de variables se lo explica en la fase del modelado, porque en esta fase de la minería aún no se encuentran categorizados los datos.

3.2.4 Verificación de los datos

Luego de haber hecho una exploración de los datos se encontraron algunas novedades en las siguientes variables:

Estado_civil.- Se encontraron datos vacíos y el estado otros que es inusual.

Edad.- Hay datos vacíos y edades menores a 16 años.

3.3 Fase III. Preparación de datos

En esta fase se analizará en cada variable aquellos datos erróneos, faltantes y se obtendrá el conjunto de datos finales para el modelo.

3.3.1 Selección de datos

Para la investigación se tienen 3 fuentes de datos:

- Entorno virtual de aprendizaje (EVA)
- Servidor de gestión Ebooks (Grammata)
- Académico (Syllabus)

Las variables de los 2 primeros sistemas poseen diferentes tipos de medición (número de interacciones y número de aplicaciones) es por ello que se ha visto necesario trabajar con dos conjuntos de datos que se detallan a continuación:

Interacción en el EVA

Se unirán los datos del EVA y del académico para la obtención de sus datos personales, las variables a utilizar son las siguientes:

Tabla 22 Interacción en el EVA

Interacción EVA	
Variables	Descripción
idnumber	Número de cédula del estudiante. Rango calculado mediante la fecha de nacimiento
edad	otorgada.
genero	Clasificación del género.
estado_civil	Clasificación del estado civil.
module	Actividades.
accion	Acciones.

Fuente: Elaboración Propia

En cuanto a las variables de module y acciones especificadas en la tabla del [Anexo 1 – D], las actividades y acciones que no se usarán son los siguientes:

Tabla 23 Evaluaciones en el Eva

Actividades	Acción
QUIZ	attempt closeAttemp conAttemp evaluacion

view
viewAll

Fuente: Elaboración propia

La actividad de evaluaciones (quiz) no se la tomará en cuenta porque es una actividad obligatoria que todos los estudiantes cumplen para poderse presentar el examen presencial y por ende no genera ninguna característica diferenciadora. Cabe indicar que esta decisión se dio por cuanto en ninguno de los cursos existía un cuestionario adicional al obligatorio.

Tabla 24 Blog en el EVA

Actividades	Acción
BLOG	View Add

Fuente: Elaboración propia

En cuanto a la actividad del blog no se la tomará en cuenta porque no registra ningún dato. Las actividades del chat (chat), foro (forums), mensajes (message) y recursos (resource) serán utilizados pero no todas sus acciones como se lo visualiza a continuación:

Tabla 25 Variables no usadas para la interacción en el EVA

Interacción EVA	
Actividad	Acción
Forums	Add (añadir foro)
	addDiss (añadir discusión)
	deleDissc (borrar discusión)
	delPost (borrar post)
	mark (marcar un foro)
	Search (buscar foros)
	update (subir al foro)
Message	updatedPost (subir un post al foro)
	userReport (ver reportes de foros)
	blockContac (bloquear un contacto)
	History (historial de mensaje)
Resource	removeContact (quitar contactos)
	unblockContact (desbloquear un contacto)
Chat	viewAll (ver un listado de todos los recursos)
	historial (historial del chat actual)
	reporte (reporte de todos los chat)
	View (ver la actividad planteada)
	viewAll (ver los chat de todas las asignaturas)

Fuente: Elaboración propia

En cuanto al foro no se tomará en cuenta las acciones de add, addDiss, deleDisc, delPost, mark, search, update, updatedPost, userReport porque son acciones que el estudiante registrar mayormente vacías y además en el planteamiento que la universidad tiene la

acción principal es la de añadir un post (**addpost**) en el foro, lo cual significa la continua interacción del estudiante en el foro planteado por el docente.

En la actividad de mensajes no se usarán las acciones blockContac, history, removeContact, unblockContact ya que son acciones muy inusuales en el estudiante y se encuentra vacío en su mayoría en este registro, la actividad principal es la de escribir un mensaje (**messageWrite**) entre compañeros o docentes.

Con respecto a la actividad de recursos consta de dos acciones de las cuales ver todos los recursos (viewAll) no se usará porque es un listado de todos los recursos de las asignaturas la acción que se usará es la de ver recursos (**resourcesView**) aquí se guarda los registros de cada recurso que ve.

Por último, en la actividad del chat no se usarán las acciones de historial, reporte, view, viewAll ya que no brinda la información de cuantas veces interactúa en el chat, esta información la brinda solamente la acción de (**chatTalk**) que es la que se usará.

La selección de estos datos se la hizo por medio de las consultas sql que se encuentran en el [Anexo 4 - A].

Servidor Gestión Ebooks

De las vistas que se analizaron en las tablas 14, 15, 16,17 las variables a utilizarse son:

Tabla 26 Variables interacción Ebooks

Interacción Ebooks	
Variables	Descripción
cedula	Cédula del estudiante
nombre	Nombre de la aplicación que utiliza el estudiantes

Fuente: Elaboración propia

El resto de variables se las uso para hacer las consultas en la base de datos como se lo observa en el [Anexo 4 - B]

Se puede observar en los datos que existen múltiples nombres de aplicaciones por ello se hizo una clasificación de aplicaciones, juegos y descargas según la categoría que presenta la tienda google play como se visualiza en la figura 23.

3.3.2 Limpieza de datos

Para saber qué hacer en la limpieza de datos se toma en cuenta lo que menciona (Pérez, 2014):

- Comprender la posible causa de la inconsistencia de los datos los posibles casos a presentarse pueden ser:

- Valores no existentes.
- Datos incompletos.
- Caracteres inválidos.

(Pérez, 2014) menciona que luego de haber comprendido que tipo de problema tienen los datos se puede hacer lo siguiente para la limpieza de datos:

- Ignorar los datos.
- Eliminar el campo.
- Filtrar la fila.
- Reemplazo de caracteres por letras.

Al analizar la variable **estado_civil** se encontraron datos vacíos y con estado otros, como el número de muestra que se requiere es de 1468 y se tiene 1630 estudiantes se procedió a eliminar estos registros porque además en la **edad** se encontraron datos menores de 10 años de esta forma se realizó la limpieza de estos datos.

En el estado civil se encontraba la opción Union Libre separado por tanto se procedió a unir las letras para evitar problemas futuros.

En cuanto a la limpieza del resto de variables se la realizará en la fase IV donde se desarrolla paso a paso la construcción del modelo.

3.3.3 Construcción e Integración de Datos

Es importante tomar en cuenta lo que se mencionó en la fase de selección de datos, donde se indicó la integración de las variables del EVA y Syllabus proceso que se llevó a cabo en un archivo excel. Este es el conjunto final de interacción en el EVA consta ya con la limpieza de datos antes mencionada en la siguiente tabla se observa el conjunto final:

Cedula	Edad_categ	Genero	EstadoCivil	Centro	ChatTalk	ForumAddPost	MessageWrite	ResourceView
:0101523009:	52_57	MASCULINO	Soltero	HUAQUILLAS	0	5	0	22
:0102005527:	52_57	FEMENINO	Casado	CUENCA	42	6	11	60
:0102086220:	46_51	FEMENINO	Soltero	CUENCA	0	4	0	26
:0102098027:	46_51	FEMENINO	Divorciado	PAUTE	15	9	9	163
:0102134103:	46_51	MASCULINO	Soltero	DURAN	0	1	4	14
:0102323854:	34_39	MASCULINO	Casado	CUENCA	34	14	34	102
:0102375383:	34_39	MASCULINO	Casado	CUENCA	50	8	1	140
:0102431236:	34_39	MASCULINO	Divorciado	CUENCA	4	11	230	9
:0102501152:	34_39	MASCULINO	Divorciado	CUENCA	8	7	5	35
:0102859907:	34_39	MASCULINO	Casado	CUENCA	0	3	3	20
:0102940665:	34_39	MASCULINO	Casado	CUENCA	29	4	25	105
:0102961646:	34_39	MASCULINO	Soltero	CUENCA	0	0	0	2
:0103173035:	40_45	FEMENINO	Casado	CUENCA	0	10	15	115
:0103311841:	28_33	FEMENINO	Casado	CUENCA	29	5	2	11
:0103441580:	40_45	FEMENINO	Soltero	IBARRA	35	7	1	79
:0103443453:	28_33	MASCULINO	Casado	SANTA ISABEL	10	7	0	9
:0103666202:	40_45	FEMENINO	Casado	CUENCA	0	8	20	130
:0103690145:	40_45	MASCULINO	Casado	CUENCA	4	4	3	41
:0103691911:	28_33	MASCULINO	Casado	CUENCA	0	5	0	7
:0103714077:	28_33	MASCULINO	Soltero	CUENCA	3	1	0	41
:0103737599:	40_45	MASCULINO	Soltero	QUITO	63	13	30	89
:0103770079:	28_33	MASCULINO	Casado	QUITOSANRAFA...	10	9	0	47

Figura 36 Conjunto de datos interacción EVA
Fuente: Elaboración personal

Luego de haber realizado la limpieza quedó un total de 1600 estudiantes para la muestra de la investigación.

En cuanto al conjunto de datos de la interacción en la Tableta digital, se realizó la clasificación como se indica en la figura 23 obteniendo el siguiente conjunto de datos:

Cedula	Edad	Genero	EstadoCivil	aplicacione																
:0102098027:		51 FEMENINO	Divorciado	0	2	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0
:0102134105:		50 MASCULINO	Soltero	0	0	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0
:0102323854:		38 MASCULINO	Casado	0	2	0	1	0	0	4	1	0	0	1	0	0	0	0	0	0
:0102375383:		38 MASCULINO	Casado	0	2	0	1	1	0	0	1	0	0	1	0	0	0	0	0	1
:0102431236:		36 MASCULINO	Divorciado	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
:0102501152:		37 MASCULINO	Divorciado	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
:0102859907:		35 MASCULINO	Casado	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
:0102940665:		34 MASCULINO	Casado	0	2	0	2	0	0	0	1	0	0	1	0	0	0	0	0	2
:0102961646:		34 MASCULINO	Soltero	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
:0103175035:		43 FEMENINO	Casado	0	1	0	1	1	0	0	1	0	0	1	1	0	0	0	0	0
:0103311841:		30 FEMENINO	Casado	0	4	0	1	1	0	0	1	0	0	1	1	0	0	0	0	0
:0103441580:		41 FEMENINO	Soltero	0	2	0	1	0	0	2	1	0	0	1	0	0	0	0	0	0
:0103443453:		33 MASCULINO	Casado	0	1	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0
:0103666202:		40 FEMENINO	Casado	0	0	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0
:0103690145:		40 MASCULINO	Casado	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
:0103691911:		32 MASCULINO	Casado	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
:0103714077:		29 MASCULINO	Soltero	0	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	1
:0103737599:		43 MASCULINO	Soltero	0	3	0	1	0	0	0	1	0	0	1	0	0	0	0	1	0
:0103770079:		32 MASCULINO	Casado	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
:0103776100:		33 MASCULINO	Casado	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
:0103880902:		38 MASCULINO	Casado	0	3	0	2	0	0	0	1	0	1	1	0	0	0	0	0	0
:0103883013:		32 MASCULINO	Divorciado	0	0	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0
:0103890877:		25 FEMENINO	Soltero	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
:0103927554:		36 FEMENINO	Soltero	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
:0103941555:		28 MASCULINO	Soltero	0	0	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0
:0104012448:		28 FEMENINO	Casado	0	0	0	1	1	0	0	1	0	0	1	0	0	0	0	0	0
:0104034061:		31 MASCULINO	Soltero	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
:0104078688:		30 MASCULINO	Soltero	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
.....	

Figura 37 Conjunto de datos Interacción en la Tableta digital

Fuente: Elaboración Propia

3.4 Fase IV. Modelado

Una vez construido los conjuntos de datos finales se procede a la creación del modelado, tomando en cuenta la técnica seleccionada en la [sección 2.2 Diseño de la solución].

3.4.1 Selección de la técnica del modelado

Para la selección correcta de la técnica de modelado se tomó en cuenta lo siguiente:

- Agrupación según características de los estudiantes.
- Adaptable al problema planteado.

Como se indicó en la [sección 2.2.3 Técnicas de minería de datos] se usará la técnica de clustering k – means.

3.4.2 Generación de la prueba de diseño

Como se mencionó anteriormente se usará la técnica del clustering k – means para la generación del modelo, es por esto que aquí se planteará el mecanismo para probar la validez de modelo.

Como lo explica (De la Fuente Fernández, 2011d) la regresión logística valida el modelo basándose en el coeficiente de verosimilitud, donde se utilizan varias pruebas estadísticas para la verificación de robustez del modelo como se visualiza a continuación:

- Verosimilitud.- esta prueba es muy importante ya que mide si un modelo se ajusta perfectamente a los datos, entre más pequeño sea el valor mejor será el ajuste (De la Fuente Fernández, 2011d).

Se debe tomar en cuenta que el resultado se basa en la resta de valores del modelo de intercepción y el modelo ajustado final.

Presenta una significancia cuando el valor de chi – cuadrado es el siguiente: ($p < 0.05$).

- Bondad de ajuste.- en esta prueba se realiza la comparación del ajuste del modelo resultante con el modelo inicial, esto determinará si el modelo es apropiado para la investigación. Se obtendrá el valor de Pearson y de la Desviación, para que el modelo se ajuste a los datos, sus valores resultantes deben ser no significativos es decir que ($p > 0,05\%$).
- Pseudo R-cuadrado.- aquí se podrá ver la versatilidad de las variables dependientes expresadas en el modelo.

En cuanto a los valores resultantes cuando sean cercanos a 1 señalarán un ajuste alto en cambio los valores cercanos a cero significan todo lo contrario.

También se obtendrá el test de R cuadrada de Nagelkerke donde se analiza el porcentaje exactitud del modelo, cuanto más alto sea el resultado de R cuadrado más demostrativo será el modelo.

Las pruebas antes mencionadas se las realizará con la obtención de los conglomerados.

3.4.3 Construcción del modelo: Entorno Virtual de Aprendizaje (EVA)

Como se mencionó en la sección [2.2.3 Técnicas de minería de datos] se eligió usar la técnica de clusterización para la obtención de patrones comportamientos basados en las acciones de los estudiantes.

Como lo menciona Mark S. Aldenderfer y Roger K. Blashfield (como se citó en Díaz de Rada, 1998) existen 4 etapas para realizar el análisis de un clúster cómo se las menciona a continuación:

- Análisis de las variables a utilizar.
- Detección de valores atípicos.
- Selección de la técnica de clasificación.
- Validación del modelo.

3.4.3.1 Análisis de las variables

Se toma en cuenta que el análisis de estas variables se realizó en la sección [3.3.1 Selección de datos] y en la [3.3.3 Construcción e Integración] donde se explicó qué variables se usarían y el porqué de su elección. El conjunto de datos finales se lo observa en la figura 35.

Se analizará las frecuencias de las variables de interacciones ya que las de datos personales fueron evaluadas en la sección [3.2.3 Exploración de datos].

Los datos que se muestran son del total de interacción de los primeros 3 meses del periodo académico en análisis.

Tabla 27 Análisis estadístico descriptivo de frecuencias

		Estadísticos			
		ChatTalk	ForumAddPost	MessageWrite	ResourceView
N	Válidos	1600	1600	1600	1600
	Perdidos	0	0	0	0
Media		18,44	6,06	6,63	61,43
Mediana		11,00	6,00	3,00	46,00
Moda		0	0	0	0 ^a
Desv. típ.		23,593	5,861	13,677	59,785
Varianza		556,615	34,354	187,059	3574,267
Mínimo		0	0	0	0
Máximo		251	136	230	778

a. Existen varias modas. Se mostrará el menor de los valores.

Fuente: Elaboración propia

Se observa que en cuanto a la media en uso de la interacciones usan 11 veces el chat, 6 veces postean en el foro, 3 veces escriben un mensaje y ven 46 recursos.

Detección de Valores Atípicos

(Perez Lopez, 2005) menciona que los clúster son muy sensibles con respecto a este tipo de datos ya que pueden cambiar drásticamente el resultado, esto se debe a que poseen un comportamiento diferente al resto, para poder identificar este tipo de datos se pueden usar un diagrama de cajas el cual es muy gráfico, permitiendo identificar dichos valores claramente como se lo visualiza a continuación:

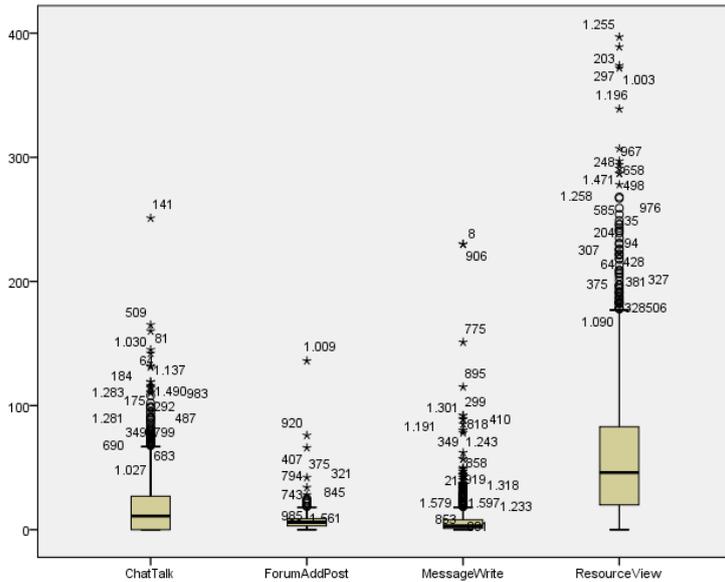


Figura 38 Valores atípicos EVA

Fuente: Elaboración propia

Para poder saber que se debe hacer con estos datos se debe tomar en cuenta lo que menciona (Perez Lopez, 2005) en identificar si su presencia es por un error que se cometió al momento de ingresar los datos o representan un comportamiento diferente el cual es importante destacar.

Se debe tener en cuenta que los datos atípicos se representan por un círculo, al momento de analizar estos datos se concluyó que es un comportamiento diferente que influyen en una gran parte de la población convirtiéndose en un conjunto importante para analizar ya que generará un patrón diferente.

3.4.3.2 Selección de la técnica de clasificación

Como se vio en la sección [2.2 Diseño de la solución] se decidió usar una tarea descriptiva con una técnica de agrupamiento, porque nos permitirá el agrupamiento del comportamiento de los estudiantes.

En cuanto a la tarea a utilizar antes de ir a la clusterización es necesario realizar un análisis factorial para ver si se necesita la reducción de variables con el fin de tener grupos homogéneos sin alterar información relevante, pero este proceso se lo realiza luego de evaluar que el conjunto de datos cumplan con lo que menciona (De la Fuente Fernández, 2011b):

En el análisis factorial se debe implementar una matriz de correlación, la misma que indicará si es necesario ejecutar dicho análisis en base al grado de asociación entre variables. Se

deber recordar que esta matriz es definida por el coeficiente de correlación de Pearson y además tomar en cuentas los siguientes indicadores para su correcto análisis:

Como lo indica (De la Fuente Fernández, 2011c) el valor de la determinante debe ser cercano a cero pero no necesariamente cero.

Prueba KMO (Kaiser-Meyer-Olkin).- (Quintín & Paz, 2007) indica que es un índice que permite la comparación de magnitudes de los coeficientes de correlación observados con las magnitudes de los coeficientes de correlación parcial.

Este índice utiliza la siguiente tabla de valores para decidir si es recomendable o no hacer el análisis factorial:

$0,90 < KMO \leq 1$	Muy bueno
$0,80 < KMO \leq 0,90$	Satisfactorio
$0,70 < KMO \leq 0,80$	Mediano
$0,60 < KMO \leq 0,70$	Mediocre
$0,50 < KMO \leq 0,60$	Bajo
$KMO \leq 0,50$	Inaceptable

Figura 39 Tabla de valores KMO

Fuente: (Quintín & Paz, 2007)

Cuando los resultados de KMO son igual o menor a 0,70 observamos que es un estado mediocre para realizar el análisis factorial por lo cual no es recomendable, pero si superan 0,70 se debe realizar el análisis factorial ya que significa que existe una correlación alta.

Porcentaje de varianza.- (De la Fuente Fernández, 2011c) explica que este método se basa en “tomar como número de factores el número mínimo necesario para que el porcentaje acumulado de la varianza explicado alcance un nivel satisfactorio (75%, 80%) ” (p.13). Tomando en cuenta que si el valor es menor al señalado no es recomendable realizar el análisis factorial.

Prueba de esfericidad de Bartlett.- indica si es prudente la aplicabilidad del análisis factorial lo que significa que debe existir una correlación significativa entre las variables cuando el resultado sea menor de 0.05 se debe realizar un análisis factorial pero si es mayor no se debe realizar este análisis.

Luego de haber revisado el tipo de pruebas necesarias para decidir si se debe realizar el análisis factorial a continuación se procede a ejecutar dichas pruebas:

La primera prueba es la matriz de correlación donde se toma en cuenta el resultado de la determinante como se lo ve a continuación:

a. Determinante = ,736

Figura 40 Determinante de la matriz de correlaciones EVA

Fuente: Elaboración propia

Se puede observar que en cuanto al resultado del determinante no es cercano a cero convirtiéndose en la primera prueba que indica que no se debe realizar el análisis factorial

Medida de adecuación muestral de Kaiser-Meyer-Olkin.		,653
Prueba de esfericidad de	Chi-cuadrado aproximado	489,809
Bartlett	gl	6
	Sig.	,000

Figura 41 Pruebas KMO y Bartlett EVA

Fuente: Elaboración propia

En cuanto a los resultados de la prueba de KMO se observa que se tiene un valor mediocre para realizar el análisis factorial, aunque en la prueba de Bartlett nos dice que tiene un grado de significancia, se analizará la varianza para tomar una decisión.

Componente	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	1,703	42,575	42,575	1,703	42,575	42,575
2	,931	23,266	65,841			
3	,733	18,315	84,156			
4	,634	15,844	100,000			

Método de extracción: Análisis de Componentes principales.

Figura 42 Varianza EVA

Fuente: Elaboración propia

Finalmente el valor de la varianza es muy bajo (42,575%) comparado con el requerimiento que se analizó anteriormente donde el valor debe ser mayor que el 70%.

Luego de evaluar cada una de las pruebas se llegó a la conclusión que no se debe realizar el análisis factorial, debido a que no cumplen las pruebas con los valores necesarios para realizar este análisis.

Por ende se procede a realizar el análisis de clusterización tanto en la herramienta SPSS y WEKA, pero teniendo en cuenta lo que menciona (De la Fuente Fernández, 2011a) para una correcta aplicación de clusterización:

- Ausencia de correlación entre las variables (validado con la pruebas de la determinante, kmo, barlet y la varianza).
- Número reducido de variables. (validado en la sección 3.3.2 Limpieza de datos y en el 3.3.3 Construcción e integración de datos).

- Homogeneidad en la medida de las variables (validado en la sección 3.3.2 Limpieza de datos y en el 3.3.3 Construcción e integración de datos).

El conjunto de datos cumple con las condiciones mencionadas, permitiendo la implementación de clusterización.

3.4.3.3 Clusterización interacción en EVA mediante WEKA

Para implementar el algoritmo K-means se debe realizar las siguientes configuraciones:

En el parámetro displayStDevs se dejará el valor por defecto que es false (falso) si en caso se quisiera saber los valores de desviación estándar se colocaría True, pero en nuestro caso quedará el valor por defecto, en el parámetro de distanceFunction se elige el tipo de distancia en este caso se quedará la más utilizada que es EuclideanDistance.

En cuanto a dontReplaceMissingValue se dejará su valor por defecto, en el parámetro de numCluster se definirá luego de haber realizado las pruebas con 2, 3 y 4 clúster como se visualizará a continuación:

```
-----
Number of iterations: 12
Within cluster sum of squared errors: 31.95037892225445
Missing values globally replaced with mean/mode

Cluster centroids:
Attribute      Full Data      Cluster#
                (1600)         0           1
                (399)         (1201)
=====
ChatTalk       18.3756        33.9574     13.199
ForumAddPost   6.0269         9.0226      5.0316
MessageWrite   6.5419         9.995       5.3947
ResourceView   60.5119        133.7644    36.1757

Time taken to build model (full training data) : 0.04 seconds

=== Model and evaluation on training set ===

Clustered Instances

0      399 ( 25%)
1     1201 ( 75%)
```

Figura 43 Clusterización numCluster 2
Fuente: Elaboración propia

```

Number of iterations: 17
Within cluster sum of squared errors: 25.327575254662513
Missing values globally replaced with mean/mode

Cluster centroids:
Attribute      Full Data      Cluster#
                (1600)         0           1           2
                (1600)         (309)       (1027)      (264)
=====
ChatTalk       18.3756        51.9741     7.6018      20.9621
ForumAddPost   6.0269         9.5178      4.3963      8.2841
MessageWrite   6.5419         8.9644      5.073       9.4205
ResourceView   60.5119        68.5631     34.1723     153.553

Time taken to build model (full training data) : 0.08 seconds

=== Model and evaluation on training set ===

Clustered Instances
0      309 ( 19%)
1     1027 ( 64%)
2      264 ( 17%)

```

Figura 44 Clusterización numCluster 3
Fuente: Elaboración propia

```

Number of iterations: 24
Within cluster sum of squared errors: 20.48627785009708
Missing values globally replaced with mean/mode

Cluster centroids:
Attribute      Full Data      Cluster#
                (1600)         0           1           2           3
                (1600)         (211)       (424)       (94)        (871)
=====
ChatTalk       18.3756        62.1943     15.4575     31.4043     7.775
ForumAddPost   6.0269         10.0758     7.0142      9.4894      4.1917
MessageWrite   6.5419         8.6114      7.0142     12.617      5.155
ResourceView   60.5119        66.3934     95.1344     210.0213    26.0976

Time taken to build model (full training data) : 0.13 seconds

=== Model and evaluation on training set ===

Clustered Instances
0      211 ( 13%)
1     424 ( 27%)
2      94 ( 6%)
3     871 ( 54%)

```

Figura 45 Clusterización numCluster 4
Fuente: Elaboración propia

Se puede observar que en comparación la figura 42, 43 y 44 en cuanto a la distribución se ve mejor y más homogénea en la figura 43 ya que la diferencia de uso entre cada clúster es clara en cambio al usar 4 clúster los grupos tienen valores parecidos entre sí. Además el valor de error cuadrático es medio en comparación a la figura 43 y 44, por esta razón se usarán 3 clúster.

3.4.3.4 Clusterización interacción en EVA mediante SPSS

Luego de la definición del parámetro numcluster se procede a ejecutar el mismo archivo en SPSS para poder evaluar en que herramienta se guarda el valor del clúster al cual

pertenece cada estudiante. Esto permitirá relacionarlos con los datos de interacción en la Tableta digital.

Centros de los conglomerados finales

	Conglomerado		
	1	2	3
ChatTalk	27	11	33
ForumAddPost	8	4	9
MessageWrite	7	5	12
ResourceView	84	25	185

Número de casos en cada conglomerado

Conglomerado	1	559,000
	2	892,000
	3	149,000
Válidos		1600,000
Perdidos		,000

Figura 46 Clusterización interacción EVA SPSS

Fuente: Elaboración propia

Se puede visualizar que los grupos de clusterización son muy parecidos a los que se obtuvo en WEKA, con la diferencia que en SPSS permite guardar el número de clúster al que pertenece cada uno de los estudiantes.

	Cedula	Edad_categorizada	Genero	EstadoCivil	Centro	ChatTalk	ForumAddPost	MessageWrite	ResourceView	InteraccionEva
1	:0101523009:	51_56	MASCULINO	Soltero	HUAQUILLAS	0	5	0	22	2
2	:0102005527:	51_56	FEMENINO	Casado	CUENCA	42	6	11	60	1
3	:0102086220:	45_50	FEMENINO	Soltero	CUENCA	0	4	0	26	2
4	:0102098027:	51_56	FEMENINO	Divorciado	PAUTE	15	9	9	163	3
5	:0102134103:	45_50	MASCULINO	Soltero	DURAN	0	1	4	14	2
6	:0102323854:	34_39	MASCULINO	Casado	CUENCA	34	14	34	102	1
7	:0102375383:	34_39	MASCULINO	Casado	CUENCA	50	8	1	140	3
8	:0102431236:	34_39	MASCULINO	Divorciado	CUENCA	4	11	230	9	2
9	:0102501152:	34_39	MASCULINO	Divorciado	CUENCA	8	7	5	35	2
10	:0102859907:	34_39	MASCULINO	Casado	CUENCA	0	3	3	20	2
11	:0102940665:	34_39	MASCULINO	Casado	CUENCA	29	4	25	105	1
12	:0102961646:	34_39	MASCULINO	Soltero	CUENCA	0	0	0	2	2
13	:0103173035:	40_45	FEMENINO	Casado	CUENCA	0	10	15	115	1
14	:0103311841:	28_33	FEMENINO	Casado	CUENCA	29	5	2	11	2
15	:0103441580:	40_45	FEMENINO	Soltero	IBARRA	35	7	1	79	1
16	:0103443453:	28_33	MASCULINO	Casado	SANTA ISABEL	10	7	0	9	2
17	:0103666202:	40_45	FEMENINO	Casado	CUENCA	0	8	20	130	1
18	:0103690145:	40_45	MASCULINO	Casado	CUENCA	4	4	3	41	2
19	:0103691911:	28_33	MASCULINO	Casado	CUENCA	0	5	0	7	2
20	:0103714077:	28_33	MASCULINO	Soltero	CUENCA	3	1	0	41	2
21	:0103737599:	40_45	MASCULINO	Soltero	QUITO	63	13	30	89	1

Figura 47 Asignación del número de cluster

Fuente: Elaboración propia

Es por esta característica especial de SPSS, que se decidió usar esta herramienta para encontrar el resto de patrones de comportamiento.

En conclusión los patrones de interacción según los clúster obtenidos se los categoriza de la siguiente forma:

Tabla 28 Categorización de la interacción en el EVA

Centros de los conglomerados finales	Interacción		
	Alta	Media	Baja
ChatTalk	33	27	11
ForumAddPost	9	8	4
MessageWrite	12	7	5
ResourceView	185	84	25

Fuente: Elaboración propia

De manera más gráfica los patrones de comportamiento en el EVA quedarían de la siguiente forma:

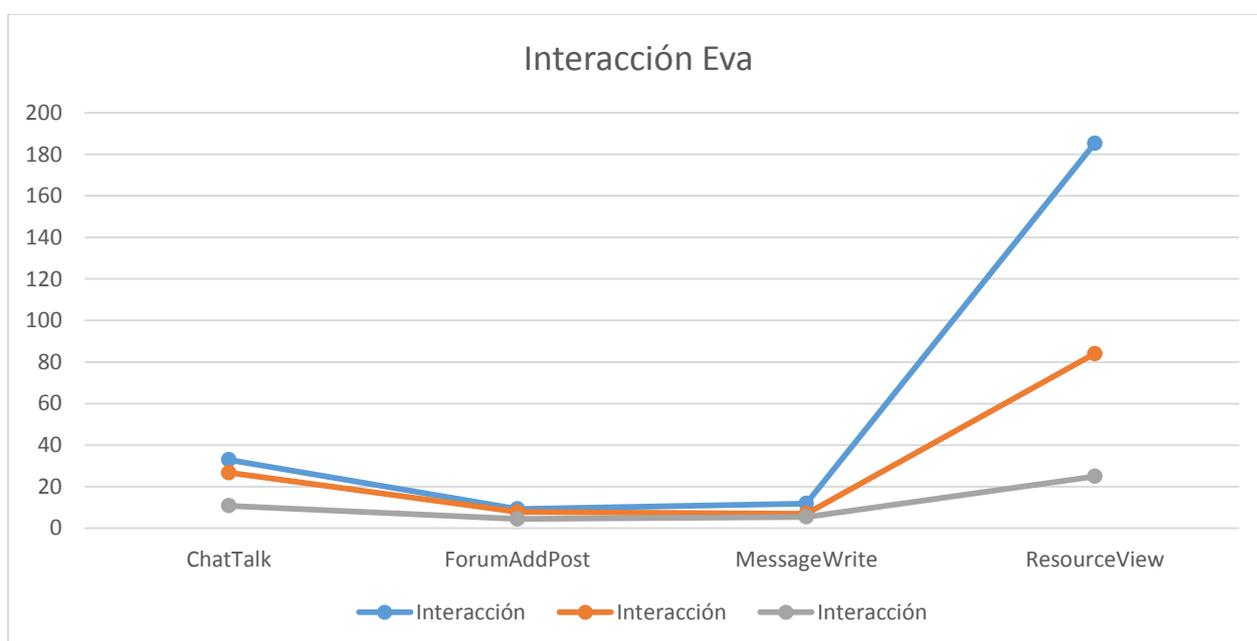


Figura 48 Patrones de comportamiento en el EVA

Fuente: Elaboración propia

Es importante conocer las características de cada uno de los clúster como se lo ve a continuación:

Interacción Alta.- son estudiantes que mayormente presentan una interacción más alta en cuanto la interacción en el chat y en vista de recursos.

Interacción media.- su característica es la de interacción media en todas las actividades de chat, postear en el foro, escribir mensajes y ver recursos.

Interacción baja.- Son estudiantes que mayormente no tienen una participación en el chat, foro, mensajes aunque en la vista de recursos si presentan una interactividad mayor a cero. Luego de obtener los patrones de comportamiento de interacción en el EVA es importante saber cómo se encuentran distribuidos en la muestra:

Tabla 29 Distribución de patrones de comportamiento

Número de casos en cada

conglomerado		
Interacción	Media	559
	Baja	892
	Alta	149
	TOTAL	1600

Fuente: Elaboración propia

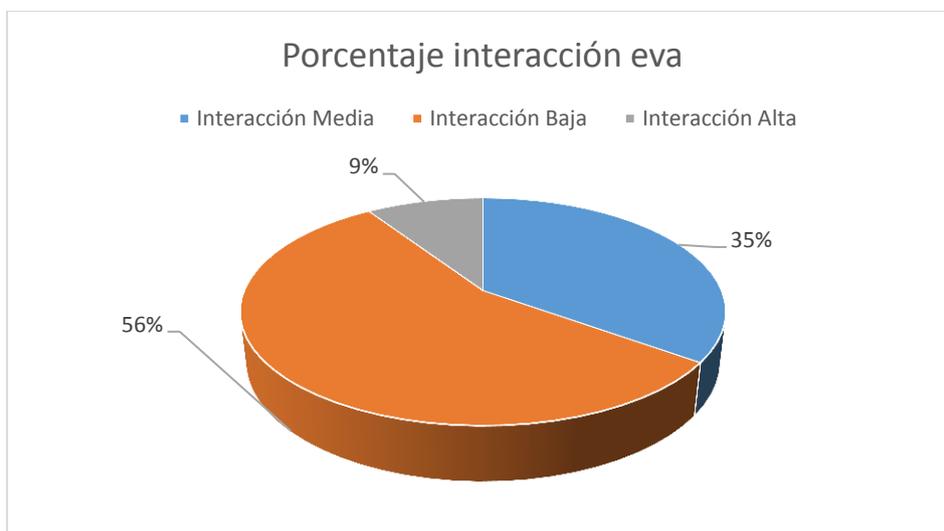


Figura 49 Porcentaje interacción en el EVA

Fuente: Elaboración propia

La interacción baja es predominante en la muestra con un 56% mientras la interacción alta se da solo en un 9%.

Es necesario saber cómo se encuentra distribuidos estos patrones en cuanto: edad, estado civil y género, esto se realizará mediante la opción de tablas de contingencia en SPSS.

Los resultados de esta distribución se encuentran en el [Anexo 7 – A: Género, estado_civil, edad], encontrando las siguientes características principales:

- La interacción alta ocupa un 9% de la muestra total, en su mayoría son personas de género masculino con estado civil soltero entre 28 a 33 años.
- La interacción media en cambio ocupa un 35% de la muestra total, en mayor porcentaje las personas son de género femenino, estado civil soltero entre 22 a 27 años.
- La interacción baja ocupa la mayor porción de la muestra con un 56%, la misma que en su mayoría corresponde al género masculino con estado civil soltero entre 16 a 21 años.

Por tanto se destacan los siguientes patrones de comportamiento:

- La población de menor interacción son los hombres entre edades de 16 a 21 años.
- La población que aunque no posee una interacción alta pero registra una interacción considerable se destacan las mujeres entre 22 a 27 años.
- Los hombres de 28 a 33 años presentan una interacción alta en un bajo porcentaje.

- Un patrón general es que las personas poseen mayormente un estado civil soltero ya sea que presenten un interacción alta, media o baja.

3.4.4 Construcción del modelo: interacción en la Tableta digital mediante SPSS

Luego de haber realizado la clusterización de la interacción en el EVA en las herramientas WEKA y SPSS se vio necesario trabajar el conjunto de datos en SPSS ya que permite guardar el clúster correspondiente a cada estudiante para luego relacionar la interacción del EVA y la Tableta digital.

Es necesario tomar en cuenta que en los datos de la interacción de la Tableta digital se realizó la categorización de las aplicaciones y juegos que se puede observar explicada a detalle en el [Anexo 6], para luego hacer un análisis estadístico de frecuencias donde se obtuvieron los siguientes resultados:

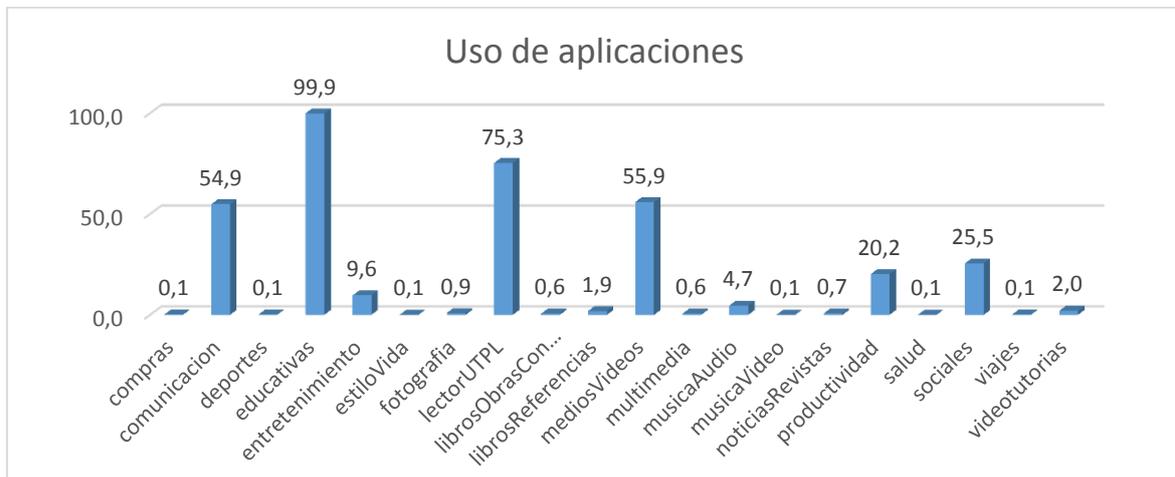


Figura 50 Uso de aplicaciones en la Tableta digital digital
Fuente: Elaboración propia

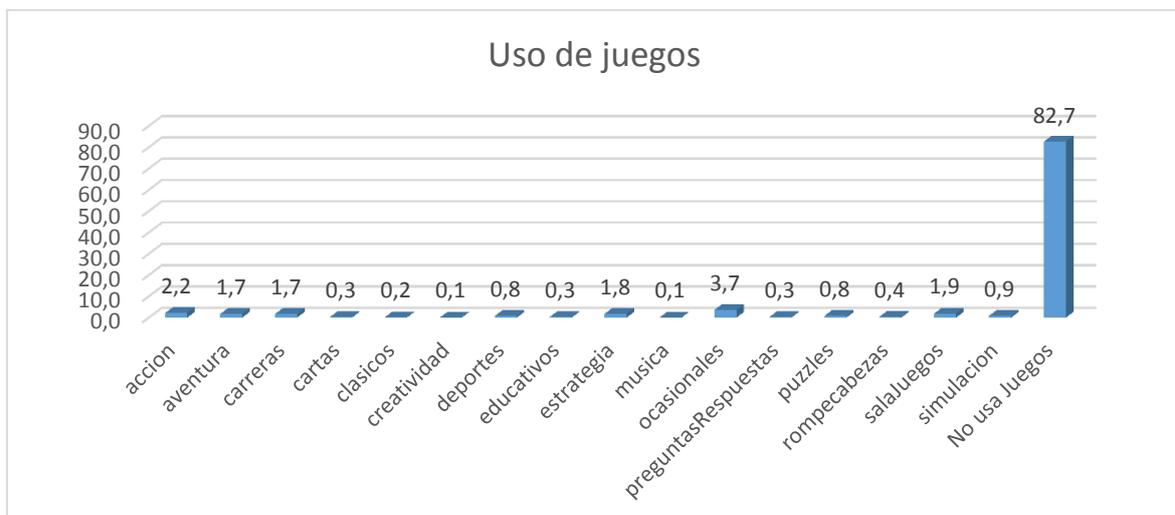


Figura 51 Uso de juegos en la Tableta digital digital
Fuente: Elaboración propia

Luego de haber observado los resultados de uso de las aplicaciones y juegos se decidió utilizar las variables de mayor uso que son:

Aplicaciones_comunicacion con un 54.9%.

Aplicaciones_educativas con un 99.9%

Aplicaciones_lectorUTPL con un 75.3%

Aplicaciones_mediosVideos con un 55.9%

Aplicaciones_productividad con un 20.2%

Aplicaciones_sociales con un 25.5%

Se encontró una peculiaridad en las variable de aplicaciones_educativas el 99,9% usa está aplicación, la variable aplicaciones_lectorUTPL utiliza un 75,3% por ende no se las tomará en cuenta ya que significan características generales en la muestra y no marcan una diferencia en los clúster, tomando en cuenta que la aplicaciones_lectorUTPL es indispensable para que el estudiante acceda a todo su material bibliográfico.

En cuanto a los juegos no se tomó en cuenta estas variables debido a que el porcentaje más elevado de uso es de 3.7%, resultando un valor extremadamente bajo convirtiéndose en una variable no homogénea dificultando su comparación.

Al igual que en la construcción del modelo de interacción en el EVA se debe evaluar si es necesario implementar el análisis factorial, mediante las siguientes pruebas:

Como resultado de la determinante de la matriz de correlación se obtuvo el valor que se ve a continuación:

Determinante = 0,290

La determinante no es cercana a cero esto indica que no se debe realizar el análisis factorial, pero se realizan más pruebas para corroborar esta conclusión.

Medida de adecuación muestral de Kaiser-Meyer-Olkin.		,563
Prueba de esfericidad de	Chi-cuadrado aproximado	1974,268
Bartlett	gl	10
	Sig.	,000

Figura 52 Pruebas KMO y Bartlett Tableta digital digital

Fuente: Elaboración propia

En cuanto a los resultados de la prueba de KMO se observa que se tiene un valor bajo (0.563) el cual indica que no se debe realizar el análisis factorial, pero en la prueba de

Bartlett nos indica un grado de significancia(0,000), para aclarar esta contradicción y tomar una decisión, se analizará la varianza.

Varianza total explicada

Componente	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	2,542	50,839	50,839	2,542	50,839	50,839
2	,902	18,046	68,885			
3	,705	14,098	82,983			
4	,463	9,256	92,239			
5	,388	7,761	100,000			

Método de extracción: Análisis de Componentes principales.

Figura 53 Varianza Tableta digital

Fuente: Elaboración propia

En vista que el el valor obtenido de la varianza es muy bajo (50,839%) con respecto al valor mínimo (70%) para realizar un análisis factorial, no se lo debe realizar debido a que no cumplen las pruebas con los valores necesarios.

Se realiza el análisis de clusterización a continuación:

Clusterización interacción en la Tableta digital digital mediante SPSS

Este conjunto utilizará 3 clúster como se lo hizo en la clusterización de interacción en el EVA.

Centros de los conglomerados finales

	Conglomerado		
	1	2	3
aplicaciones_comunicacion	0	3	2
aplicaciones_mediosVideos	0	1	1
aplicaciones_productividad	0	0	2
aplicaciones_sociales	0	0	0

Número de casos en cada conglomerado

Conglomerado	1	1106,000
	2	331,000
	3	163,000
Válidos		1600,000
Perdidos		,000

Figura 54 Clusterización interacción Tableta digital digital SPSS

Fuente: Elaboración propia

Se procede a guardar el número de clúster en cada uno de los registros de los estudiantes como se visualiza en la siguiente figura:

Cedula	Edad_categ	Genero	Estado_Civil	Centro	Interaccion Eva	aplicaciones_comunicacion	aplicaciones_mediosVideos	aplicaciones_productividad	aplicaciones_sociales	InteraccionTableta
:0101523009:	52_57	MASCULINO	Soltero	HUAQUILLAS	2	0	0	0	0	1
:0102005527:	52_57	FEMENINO	Casado	CUENCA	1	4	1	1	1	2
:0102086220:	46_51	FEMENINO	Soltero	CUENCA	2	0	0	0	0	1
:0102098027:	46_51	FEMENINO	Divorciado	PAUTE	3	2	1	0	0	2
:0102134103:	46_51	MASCULINO	Soltero	DURAN	2	0	1	0	0	1
:0102323854:	34_39	MASCULINO	Casado	CUENCA	1	2	1	0	0	2
:0102375383:	34_39	MASCULINO	Casado	CUENCA	3	2	1	1	0	3
:0102431236:	34_39	MASCULINO	Divorciado	CUENCA	2	0	0	0	0	1
:0102501152:	34_39	MASCULINO	Divorciado	CUENCA	2	0	0	0	0	1
:0102859907:	34_39	MASCULINO	Casado	CUENCA	2	0	0	0	0	1
:0102940665:	34_39	MASCULINO	Casado	CUENCA	1	2	1	2	0	3
:0102961646:	34_39	MASCULINO	Soltero	CUENCA	2	0	0	0	0	1
:0103173035:	40_45	FEMENINO	Casado	CUENCA	1	1	1	0	0	1
:0103311841:	28_33	FEMENINO	Casado	CUENCA	2	4	1	0	0	2
:0103441580:	40_45	FEMENINO	Soltero	IBARRA	1	2	1	0	1	2
:0103443453:	28_33	MASCULINO	Casado	SANTA ISABEL	2	1	1	0	0	1
:0103666202:	40_45	FEMENINO	Casado	CUENCA	1	0	1	0	0	1
:0103690145:	40_45	MASCULINO	Casado	CUENCA	2	0	0	0	0	1
:0103691911:	28_33	MASCULINO	Casado	CUENCA	2	0	0	0	0	1
:0103714077:	28_33	MASCULINO	Soltero	CUENCA	2	1	0	1	1	1
:0103737599:	40_45	MASCULINO	Soltero	QUITO	1	3	1	0	1	2
:0103770070:	28_33	MASCULINO	Casado	QUITO	2	0	0	0	0	1

Figura 55 Registro de interacción en las Tableta digital digital

Fuente: Elaboración propia

Luego de la ejecución de clúster en SPSS se obtuvo la siguiente clasificación de los grupos de interacción en la Tableta digital digital:

Tabla 30 Categorización de la interacción en la Tableta digital digital

	Centros de los conglomerados finales		
	Básico	Comunicativo-social	Productivo
aplicaciones_comunicacion	0	3	2
aplicaciones_mediosVideos	0	1	1
aplicaciones_productividad	0	0	2
aplicaciones_sociales	0	0	0

Fuente: Elaboración propia

La denominación de los clúster no se la hizo de tipo: bajo medio y alto, ya que la diferenciación de los clúster no se la realizó en base al número de interacciones sino en base al número y tipo de aplicaciones.

Para comprender mejor los patrones de interacción en las Tableta digital observe el siguiente gráfico:

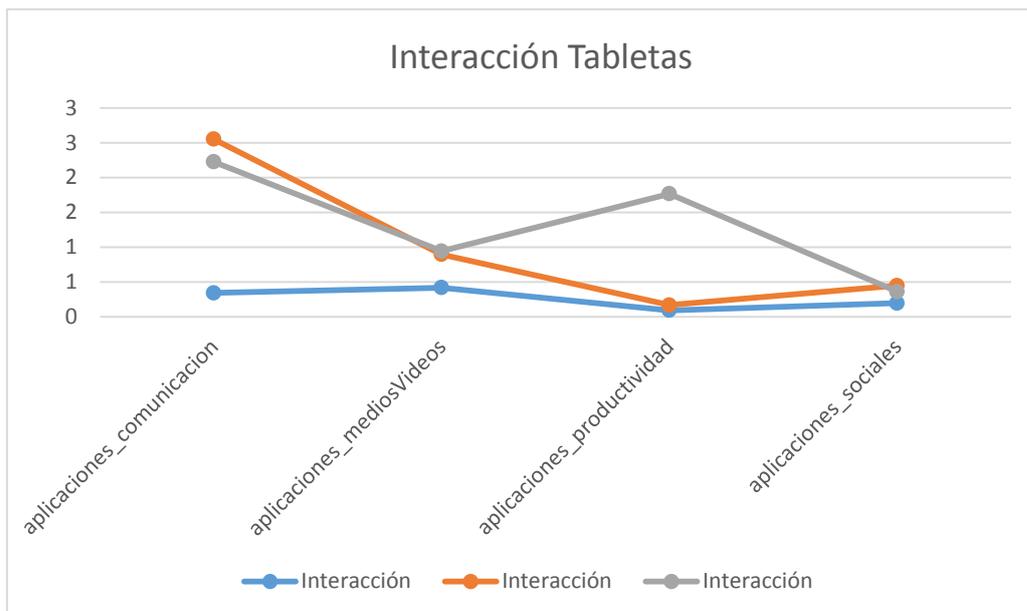


Figura 56 Patrones de comportamiento en las Tableta digital

Fuente: Elaboración propia

Se observa características principales en cada grupo de interacciones o grupo de patrones de comportamiento:

- **Interacción básica:** estudiantes que adicional a la descarga, lectura del Ebooks y acceso al EVA utilizan muy poco o nada otro tipo de aplicaciones.
- **Interacción comunicativo – social:** este grupo posee el uso más alto de todos los grupos en aplicaciones de comunicación como: correo electrónico, navegadores, Skype, line, redes sociales entre otros y un uso medio en cuanto aplicaciones de medios_videos como: youtube, vlc entre otros.
- **Interacción productiva:** este grupo posee un uso más alto de aplicaciones para producir y compartir contenido como: dropbox, herramientas de ofimática, prezzi, goggle drive además de un uso medio en aplicaciones de comunicación, sociales, medios y videos.

Es importante conocer la distribución de estos patrones de comportamiento en la muestra:



Figura 57 Porcentaje interacción Tableta digital digital

Fuente: Elaboración propia

La mayor parte de la muestra tiene una interacción básica en las Tabletas digital con un 69% y una interacción productiva solamente con un 10%, esto no significa que no usen la Tableta por qué se debe recordar que un patrón general que posee la muestra es el uso de la Tableta para leer sus ebooks y revisar el aplicativo del escritorio UTPL.

Estos patrones de comportamientos poseen otras características en relación a la edad, género y estado civil en el [Anexo 7 – B: Género, estado_civil, edad]:

- La interacción básica ocupa la mayoría de la muestra en un 69%, de este porcentaje mayormente corresponden al género masculino, estado civil soltero entre 22 a 27 años.
- La interacción comunicativo-social ocupa el 21% de la muestra, de este porcentaje tienen mayormente las siguientes características: género femenino, estado civil soltero, entre 16 a 21 años.
- En cuanto a la interacción productiva ocupa el 10% de la muestra, en su mayoría género femenino, estado civil soltero entre los 22 a 27 años.

Finalmente se realizará la comparación de la interacción del EVA con la Tableta digital como observa en la siguiente tabla:

Tabla 31 Interacción Tableta digital digital - EVA

		InteraccionEva			Total
		Alto	Medio	Bajo	
InteraccionTableta digital digital	Bajo	5%	23%	41%	69%
	Comunicativo-social	3%	8%	10%	21%
	Productivo	1%	4%	5%	10%

Fuente: Elaboración propia

Se puede observar que en todas las interacciones de la Tableta digital se tiene presente interacciones del Eva.

Existiendo varios patrones de comportamiento como los que se menciona a continuación:

- Tanto en el EVA como en la Tableta poseen mayormente una interacción baja y básica en todos sus patrones lo cual significa que casi no interactúan en el EVA y la Tableta la usan solo para descargar, leer su material digital e interactuar en el escritorio UTPL.
- En cuanto el patrón de interacción bajo en el EVA también existen otros patrones de interacción con respecto a la Tableta, un 10% de la muestra posee un patrón comunicativo – social significando un uso de aplicaciones de comunicación como: correo electrónico, navegadores, Skype, line, redes sociales entre otros y un uso medio en cuanto aplicaciones de medios_videos como: youtube, vlc entre otros. Un 5% de la muestra en cambio presenta un patrón de productividad lo que significa que usa aplicaciones para para producir y compartir contenido como: dropbox, herramientas de ofimática, prezzi, goggle drive además de un uso medio en aplicaciones de comunicación, sociales, medios y videos.
- La población que presenta un uso medio en el EVA también presenta los diferentes patrones de la interacción de la Tableta un 8% son comunicativos – sociales y un 4% productivos.
- En la interacción alta del EVA se puede dar cuenta que mayormente los estudiantes usan la Tableta solo para su característica general y tan solo un 3% son comunicativos - sociales y un 1% productivos.

3.4.4 Evaluación del modelado

Como se lo menciono en la sección [3.4.2 Generación de la prueba de diseño] se implementara las siguientes pruebas para validar la robustez del modelo.

Modelo interacción EVA

Las pruebas que se realizaran para determinar si el modelo se ajusta a los datos son la siguiente:

Tabla 32 Prueba de verosimilitud EVA

Información del ajuste del modelo				
Modelo	Criterio de ajuste del modelo	Contrastes de la razón de verosimilitud		
	-2 log verosimilitud	Chi-cuadrado	gl	Sig.

Sólo la intersección	2925,474			
Final	,068	2925,406	796	,000

Fuente: Elaboración Propia

El resultado de la prueba de verosimilitud indica un correcto ajuste de modelo ya que su nivel de significancia es menor a 0.05.

Tabla 33 Bondad de ajuste EVA

	Chi-cuadrado	gl	Sig.
Pearson	,034	2280	1,000
Desviación	,068	2280	1,000

Fuente: Elaboración propia

Igualmente e la prueba de bondad de ajuste indica una significancia mayo a 0.05 tanto en el valor de Pearson y la desviación lo que significa un correcto ajuste de los datos al modelo.

Tabla 34 Pseudo R-cuadrado

Cox y Snell	,839
Nagelkerke	1,000
McFadden	1,000

Fuente: Elaboración propia

Igualmente la prueba de R-cuadrado indica una exactitud y ajuste correcto del modelo

Modelo interacción Tableta digital digital

A continuación se realizara la prueba de verosimilitud del modelo:

Tabla 35 Prueba de verosimilitud Tableta digital digital

Modelo	Criterio de ajuste del modelo	Contrastes de la razón de verosimilitud		
	-2 log verosimilitud	Chi-cuadrado	gl	Sig.
Sólo la intersección	2604,450			
Final	,002	2604,449	26	,000

Fuente: Elaboración propia

Presenta una significancia mejor a 0,05 lo que indica un correcto ajuste del modelo.

Tabla 36 Bondad de ajuste Tableta digital digital

	Chi-cuadrado	gl	Sig.
--	--------------	----	------

Pearson	,001	152	1,000
Desviación	,002	152	1,000

Fuente: Elaboración propia

En cuanto a la bondad de ajuste tanto el valor de Pearson y la Desviación tienen una significancia mayor a 0,05 lo que indica que el modelo se ajusta a los datos.

Tabla 37 Pseudo R-cuadrado Tableta digital

Pseudo R-cuadrado	
Cox y Snell	,804
Nagelkerke	1,000
McFadden	1,000

Fuente: Elaboración propia

Finalmente la prueba de R-cuadrado demuestra un ajuste y exactitud correcta del modelo ya que sus valores son mayores al 0.5%.

Concluyendo que el modelo empleado para la búsqueda de patrones de comportamiento en las Tableta digital es correcto.

CONCLUSIONES

Con la obtención de los modelos se pudieron cumplir con los objetivos planteados al inicio del proyecto.

A continuación se explica las conclusiones del mismo:

- Al momento de analizar las experiencias de 3 universidades se pudo observar que los resultados que obtuvieron eran dirigidos al horario, periodo de conectividad, predisposición para implementar los dispositivos móviles en su aprendizaje, conocer si se encuentran familiarizados con el uso de libros electrónicos, siendo sus respuestas negativas al momento de usarlos pero porque no poseen una capacitación adecuada además de tener costos muy altos para cada estudiante. En comparación con nuestro estudio se puede determinar que los estudiantes de la UTPL usan la Tableta digital para la lectura de su material y acceso a la plataforma de la universidad, además de generar otros comportamientos mediante el uso de otras aplicaciones que el estudiante busca por iniciativa propia.
- Cuando se realizó el análisis, exploración y verificación de los datos se determinó que la mayor población era de estado civil soltero y en un leve porcentaje mayoritario son de género masculino, en cuanto a edad la población se encuentra entre 22 a 27 años igualmente en un porcentaje levemente mayor.
- En cuanto a la selección, limpieza e integración de los datos el unir los datos personales con los de las interacciones permitió la obtención de patrones específicos.
- Siempre es importante realizar un análisis factorial previo al análisis de clúster para poder determinar si existe correlación entre las variables para realizar una reducción de variables pero en el presente estudio no existió esta correlación tomando en cuenta las pruebas de la determinante, KMO, barlet, y el análisis de la varianza, por ende se procedió directamente al análisis de clúster.
- Usar la técnica de clúster mediante el algoritmo K- means en la presente investigación permitió agrupar a los estudiantes de acuerdo a características semejantes, permitiendo una descripción e interpretación de los datos más sencilla, a su vez generando una nueva variable la cual permitió la relación entre la interactividad de la Tableta digital y el EVA.
- Usar la técnica de regresión multinomial ayudó a verificar la robustez y ajuste del modelo.
- Tanto en la interacción alta y baja del EVA el grupo mayoritario son de género masculino, en cambio en la interacción media mayormente influye el género

femenino. Lo que nos indica que las que mayormente interactúan en el EVA son las mujeres de 22 a 27 años, debido al porcentaje considerable en interacción media.

- Los estudiantes a pesar de tener una interacción baja en el EVA presentan una interacción básica en la tableta lo que indica que los estudiantes realmente usan las aplicaciones que la universidad genera específicamente para ayudar en su formación académica y además generan patrones de comportamiento comunicativos – sociales y productivos mediante el uso de aplicaciones que ellos consideren necesarias.
- A pesar de que la mayoría de estudiantes presentan un interacción básica en la tableta se tienen dos patrones de comportamiento: comunicativo – social y productivo que mayormente lo generan las mujeres de 16 a 21 años el patrón comunicativo – social y el productivo mujeres de 22 a 27 años, siendo la mujeres las que generan otros patrones de comportamiento a diferencia de los hombres quienes poseen solamente el patrón de interacción básica.
- Aunque los hombres entre 16 a 21 años presentan mayormente una interacción baja en el EVA, existe un porcentaje pequeño que llama la atención porque corresponde a la interacción alta en el EVA de hombres de 28 a 33 años, a diferencia de las mujeres que solamente poseen una interacción media.
- En la muestra de estudio la mayoría tanto en la interacción del EVA y la Tableta presenta la característica general de estado civil soltero por ello esta variable no influye algún cambio en los patrones de comportamiento debido a que es el mismo en todos los grupos.

RECOMENDACIONES

- Se cree conveniente realizar un estudio de todo un ciclo ya que el presente fue de los 3 primeros meses para poder identificar si los patrones cambian en cuanto a periodo de exámenes en comparación con el periodo normal de clases, además de implementar campos en los registros de interacción de la Tableta digital como fecha, hora y nivel de satisfacción en el uso de las herramientas que posee tanto el EVA como la Tableta digital.
- Siempre realizar un análisis exploratorio y verificación de datos para evitar que se dejen valores atípicos, caracteres erróneos que cambien el resultado de los patrones de comportamiento.
- En cuanto a los datos personales se debería incluir datos socios económicos para saber si esto influye en su comportamiento.
- Usar tanto WEKA como SPSS y comparar si los grupos son parecidos para poder validar que se obtienen los datos correctos.
- Recolectar la data de las acciones en el lector y escritorio UTPL para saber si el estudiante tiene problemas o necesita otras opciones en dichas aplicaciones.

BIBLIOGRAFÍA

- Aedo, I., & Díaz, P. (Junio de 2000). *ISI*. Obtenido de http://lsi.ugr.es/~mgea/workshops/interaccion2000/trabajos/articulos/articulos/Aedo_I.pdf
- Allen Shelburne, W. (03 de 12 de 2013). *Taylor & Francis Group*. Obtenido de <http://www.tandfonline.com/doi/pdf/10.1080/14649055.2009.10766234>
- Alonso, B. (20 de Octubre de 2013). *Sistemas de protección de e-books*. Obtenido de http://www.c2com.up.pt/tm/beatriz_alonso/2013/10/-sistemas-de-proteccion-de-e-books.html
- Area, M. (01 de Septiembre de 2013). *Red Latinoamericana Portales Educativos*. Obtenido de <http://www.relpe.org/especial-del-mes/de-los-libros-de-texto-a-los-contenidos-digitales-cambio-pedagogico-o-cambio-del-modelo-de-negocio/>
- Arévalo, A., Córdon, J., & Gómez, R. (Marzo de 2011). *Biblios*. Obtenido de <http://biblios.pitt.edu/ojs/index.php/biblios/article/view/7/37>
- Baratz, C. (10 de Febrero de 2014). *Comunidad Baratz*. Obtenido de http://www.comunidadbaratz.com/blog/todos-los-formatos-de-libros-electronicos?utm_source=HS
- Barker, P. (1996). *Electronic Books: a Review and Assessment of Current Trends*.
- Cabrera Frías, Y., & Malágon Hernández, M. J. (09 de 2009). *e-Libro*. Obtenido de <http://site.ebrary.com/lib/bibliotecautplsp/detail.action?docID=10327110&p00=libro+el+electronico>
- Carlos, L. (2010). *typepad*. Obtenido de <http://luisc33.typepad.com/blog/2010/10/proyecto-de-aprendizaje-de-las-matematicas-por-parte-de-nokia-en-sur-africa-.html>
- Castro, M. (05 de Junio de 2014). *Slideshare*. Obtenido de <http://es.slideshare.net/mmmcastro/impacto-del-binomio-tecnologaglobalizacin-en-la-docencia-en-la-educacin-superior>
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., . . . (Traducción realizada por Gutiérrez, D. (2000). *dataprix*. Obtenido de <http://www.dataprix.com/es/metodolog-crisp-dm-para-miner-datos>

- Celestino, M.A., Flores, J. & Rangel, R. (2004). Rumbo a la estadística multivariante
Recuperado de
https://books.google.com.ec/books?id=KONFKQzKbV0C&printsec=frontcover&hl=es&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false
- (Chapman et al., 2007)
- Chuquiyauri, L. (12 de Septiembre de 2014). *Peru Educa*. Obtenido de
<http://educaciontic.perueduca.pe/?p=2039>
- Ciberaula. (2012). *ciberaula*. Obtenido de <http://www.ciberaula.com/empresas/blearning>
- Comunidad Movistar*. (01 de Febrero de 2012). Obtenido de
<http://comunidad.movistar.es/t5/Blog-Tablets/Acabar%C3%A1n-las-tablets-con-los-eReaders/ba-p/449711>
- Cordón García, J. A., & Arévalo, J. A. (2010). *Las políticas de adquisición de libros electrónicos en bibliotecas*. Gijón. Obtenido de
http://eprints.rclis.org/15059/1/Las_pol%C3%ADticas_de_adquisici%C3%B3n_de_libros_electr%C3%B3nicos.pdf
- De Oliveira, J., Camacho, M., & Gisbert, M. (2014). *Comunicar*. Obtenido de
www.revistacomunicar.com/verpdf.php?numero=42&articulo=42-2014-08
- Del Valle, A. G., Fernández, P., & Garrido, S. (16 de Septiembre de 2011). *Observatorio Tecnológico*. Obtenido de
<http://recursostic.educacion.es/observatorio/web/ca/equipamiento-tecnologico/hardware/1013-monografico-ebook-definicion-autoproduccion-y-testeo-de-formatos-y-dispositivos?start=2>
- E-abclearning. (2011). *e-abc*. Obtenido de <http://www.e-abclearning.com/definicion-learning>
- Edukanda. (s.f.). *Edukanda*. Obtenido de
http://www.edukanda.es/mediatecaweb/data/zip/1308/page_03.htm
- Frontlinessolver. (2015). *Frontlinessolver*. Obtenido de <http://www.solver.com/xlminer-capabilities-overview>
- García González, F. J. (2013). *Universidad de Granada*. Obtenido de
http://masteres.ugr.es/moea/pages/tfm-1213/tfm_garciagonzalezfrancisco_1/

- García Peñalvo, F. (2005). *Campus Usal*. Obtenido de http://campus.usal.es/~teoriaeducacion/rev_numero_06_2/n6_02_art_garcia_penalvo.htm
- Groenenboom, M., & Helberger, N. (11 de Septiembre de 2006). *Guía del consumidor sobre gestión de derechos digitales*. Obtenido de http://www.indicare.org/tiki-download_file.php?fileId=201
- Hand, D., Mannila, H., & Smyth, P. (2001). *Principles of Data Mining*. MIT Press.
- Hernández Orallo, J., Ramírez Quintana, M., & Ferri Ramírez, C. (2004). *Introducción a la Minería De Datos*. España: Pearson.
- Herrera Cuenca, J. (Julio de 2013). *Oberta*. Obtenido de <http://blog.obertapublishing.com/u-learning-quitandole-barreras-al-campo/>
- Herrera, S. I., Fennema, M. C., & Sanz, C. V. (2010). *Sedici*. Obtenido de http://sedici.unlp.edu.ar/bitstream/handle/10915/18313/Documento_completo.pdf?sequence=1
- IBM. (2014). *IBM*. Obtenido de http://www-01.ibm.com/support/knowledgecenter/SSEPGG_9.5.0/com.ibm.im.model.doc/c_defining_mining_tasks2.html?lang=es
- IBM. (2015). *IBM*. Obtenido de <http://www-01.ibm.com/software/analytics/SPSS/products/modeler/>
- Instituto de Tecnologías Educativas. (s.f.). *ite*. Obtenido de http://www.ite.educacion.es/formacion/materiales/157/cd/m7_1_plataformas_de_aprendizaje_en_red/conceptos_generales.html
- Isaacs, S. (2012). *Mobile Learning for Teachers in Africa and the Middle East*. Obtenido de <http://unesdoc.unesco.org/images/0021/002163/216358E.pdf>
- Kdnuggets. (2014). *Kdnuggets*. Obtenido de <http://www.kdnuggets.com/polls/2014/analytics-data-mining-data-science-methodology.html>
- Loor Díaz, L. E. (2006). *Espol*. Obtenido de <http://www.dspace.espol.edu.ec/handle/123456789/5832>

- López Hernández, F. A., & Silva Pérez, M. M. (2014). *Journals*. Obtenido de <http://journals.uoc.edu/index.php/rusc/article/viewFile/v11n1-lopez-silva/v11n1-lopez-silva-es>
- López Herrero, A. C. (2010). *Tecnoneet* . Obtenido de Atención a la diversidad: <http://diversidad.murciaeduca.es/tecnoneet/2010/docs/aclopez.pdf>
- Marques de Oliveira, S. (2012). *Docs*. Obtenido de <http://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1122&context=iatul>
- Martí, J. (13 de Febrero de 2012). *xarxatic*. Obtenido de <http://www.xarxatic.com/el-c-learning-usando-el-potencial-de-la-nube-para-el-aprendizaje/>
- Mohamed , O., El-Hussein , & Cronje, J. (s.f.). *Journals of Educational Technology & Society*. Obtenido de http://www.ifets.info/journals/13_3/3.pdf
- Moine, J. M., Haedo, A. S., & Gordillo, S. (2013). *Sedici*. Obtenido de http://sedici.unlp.edu.ar/bitstream/handle/10915/20034/Documento_completo.pdf?sequence=1
- Moravec, J. W., & Cobo, C. (2011). *Hacia una nueva ecología de la educación*. Barcelona: Col·lecció Transmedia XXI.
- Moreno Guerrero , A. J. (17 de Diciembre de 2011). *Observatorio Tecnológico*. Obtenido de <http://recursostic.educacion.es/observatorio/web/fr/cajon-de-sastre/38-cajon-de-sastre/1026-movil-learning>
- NewWweb. (2014). *newWweb*. Obtenido de <http://www.newwweb.com.mx/LMS>
- Oxford Dictionaries. (2013). *Oxford Dictionaries*. Obtenido de <http://oxforddictionaries.com/definition/english/e-book>
- Padilla, L. (2014). *Slideshare*. Obtenido de <http://es.slideshare.net/Lucypadilla2/byod-upaep-2014-prcticas-docentes-2014-upaep>
- Pérez López, C. (2007). *Minería de datos: técnicas y herramientas*. Paraninfo.
- Peréz, F., & Moscoso, P. (2005). *Revista española de Documentación Científica*. Obtenido de <http://redc.revistas.csic.es/index.php/redc/article/viewFile/389/401>
- Quintín, M., y Paz, Y. del R.(2007). Tratamiento estadístico de los datos con SPSS. (Editorial Parainfo, ed.).

- Quispe Ortega, L. (27 de Febrero de 2011). *Devian*. Obtenido de <http://www.consultora-devian.net/inicio/noticias/52-de-la-web-10-a-la-web-40>
- Ramírez Montoya, M. S. (9 de Diciembre de 2008). *redalyc*. Obtenido de <http://www.redalyc.org/pdf/688/68811230006.pdf>
- Ramón, L. (2014) *Minería de datos* (Tesis de Maestría) Universidad Nacional del Nordeste, Argentina
- Ramos, F. (2004). *Universitat Pompeu Fabra Barcelona*. Obtenido de <http://www.upf.edu/hipertextnet/numero-2/drm.html>
- Riquelme, J. C., Ruiz , R., & Gilbe, K. (2006). *Redalyc*. Obtenido de <http://www.redalyc.org/articulo.oa?id=92502902>
- Rodríguez, E. (Abril de 2013). *Coaching Tecnológico*. Obtenido de <http://www.coaching-tecnologico.com/ventajas-y-desventajas-del-libro-electronico-o-ebook/>
- Roura Redondo, M. (Junio de 2012). *Educoas*. Obtenido de http://educoas.org/portal/la_educacion_digital/147/pdf/ART_UNNED_EN.pdf
- Salamanca, M., & Sánchez, F. G. (2010). *Territorio Ebook*. Obtenido de http://www.territorioebook.com/includes/download.php?id=../recursos/1298988498_ProyectoDedos.pdf
- Salford Systems. (2015). *Salford Systems*. Obtenido de <https://www.salford-systems.com/products>
- Sas. (2014). *Sas*. Obtenido de <http://www.sas.com/enterpriseminer>
- Tapia, E. (2013). Obtenido de <http://enriquetapia.net/la-web-semantica-o-3-0/>
- Traxle, J. (2007). *Defining, Discussing and Evaluating Mobile Learning: The Moving Finger Writes and Having Write The International Review in Open and Distance Learning*.
- Traxler, J. (2005). *Academia*. Obtenido de http://www.academia.edu/2810810/Defining_mobile_learning
- Traxler, J. (Enero de 2009). *Learning in a Mobile Age*. Obtenido de https://www.academia.edu/171500/Learning_in_a_Mobile_Age

- Unesco. (2012). *Aprendizaje Móvil Para Docentes En América Latina*. Obtenido de <http://unesdoc.unesco.org/images/0021/002160/216081s.pdf>
- Unesco. (2012). *Unesco*. Obtenido de <http://unesdoc.unesco.org/images/0021/002163/216358E.pdf>
- Unesco. (2013). *Unesco*. Obtenido de <http://unesdoc.unesco.org/images/0021/002196/219637s.pdf>
- Universia. (21 de Enero de 2013). *Universia*. Obtenido de <http://noticias.universia.es/en-portada/noticia/2013/01/21/995067/5-formatos-mas-populares-ebooks.html>
- Universidad Complutense, D. (2011). *Política de adquisición de libros electrónicos en línea de la biblioteca de la Universidad Complutense de Madrid*. Madrid. Obtenido de http://www.mecd.gob.es/cultura-mecd/eu/dms/mecd/cultura-mecd/areas-cultura/bibliotecas/mc/congresobp/2010/comunicaciones/textos-completos/textosbi/C1_4_Policasadquisicionlibelectronicos.pdf
- Vallejo, D., & Tenelanda, G. (25 de 04 de 2012). *usbmed*. Obtenido de <http://web.usbmed.edu.co/usbmed/fing/v3n1/v3n1a6.pdf>
- Villar , M. (2011). *Evo I.T.* Obtenido de <http://blog.evoit.com/2011/03/nuevas-tendencias-p-learning-y-u-learning/>
- Virtual, Aula. (Noviembre de 2012). *Aula Virtual*. Obtenido de <http://aula.virtual.ucv.cl/wordpress/t-learning-television-digital-interactiva-y-educacion/>
- Weka. (2013). *Weka*. Obtenido de <http://www.cs.waikato.ac.nz/ml/weka/>
- Woodill, G. (2011). *The mobile learning edge*. Mc Graw Hill.

ANEXOS

ANEXO 1: TABLAS DEL ENTORNO VIRTUAL DE APRENDIZAJE (EVA)

Anexo 1 – A: mdl_user

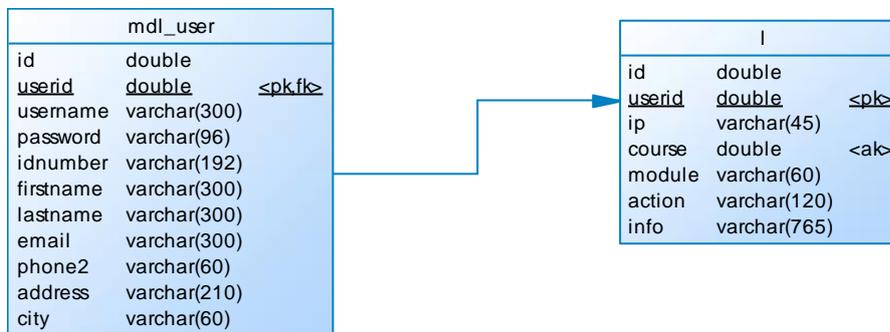
Campo	Descripción	Tipo de carácter
id	identificador de la tabla mdl_user	double
userid	Identificador del usuario	double
username	Nombre de nick del usuario	varchar
password	Clave del usuario	varchar
idnumber	Número de cédula del estudiante	varchar
firstname	Nombre del usuario	varchar
lastname	Apellido del usuario	varchar
email	Correo del usuario	varchar
phone	Teléfono del usuario	varchar
address	Dirección del usuario	varchar
city	Ciudad	varchar

Anexo 1 – B: mdl_log

Campo	Descripción	Tipo de carácter
id	Identificador de la tabla mdl_log	double
userid	Identificador del usuario	double
ip	Dirección ip de donde se conecta	varchar
course	Identificador de la materia	double

module	Tipo de actividad en el EVA	varchar
action	Tipo de acción que realiza en las actividades planteadas	varchar
info	Información adicional	varchar

Anexo 1 – C: Tablas del EVA Relacionadas



Anexo 1 – D: Variables de modules - acciones

Actividades	Acción
QUIZ	attempt
	closeAttemp
	conAttemp
	evaluacion
	view
BLOG	viewAll
	add
CHAT	historial
	reporte
	talk
	view
FORUMS	viewAll
	add
	addDiss
	addPost
	deleDisssc
	delPost
	mark
search	

	update updatedPost userReport
MESSAGE	add blockContac history removeContaC unblockContact write
RESOURCE	view viewAll

ANEXO 2: TABLA DE DATOS DEMOGRÁFICOS (SYLLABUS)

Anexo 2 – A: datos_personales

datos_demograficos	
apellidos	Text
nombre	Text
identificacion	Number
lugar	Text
modalidad	Text
mac_fecha_registro	Date
mac_numero	Date
titulacion	Text
periodo	Text
correo_personal	Text
telefonos	Number
dir_domicilio	Text
fecha_nacimiento	Date
colegio	Text
estado_matricula	Text
area	Text
centro	Text
nacionalidad	Text
genero	Text
provincia	Text
canton	Text
estado_civil	Text
discapacidad	Text
etnia	Text
edad	Number
nivel	Number

ANEXO 3: VISTAS SERVIDOR GESTIÓN EBOOKS (GRAMMATA)

Anexo 3 – A: vistas

Nombre_Vista	Descripción	Variable
biusuario	Datos personales de los estudiantes	id_usuario nomb_usuario cedula nombre apellidos
bidispositivo	Datos del dispositivo	id_dis_fisico

		num_serie
bidispositivousuario	Relación del usuario con el dispositivo	id_dis_fisico id_usuario
biaplicacion	Datos de las aplicaciones usadas en la Tableta digital	id nombre paquete sistema_operativo
bihistoriacion	Relación de las aplicaciones en los dispositivos usados	id_aplicacion id_dis_fisico

ANEXO 4: CONSULTAS SQL

Anexo 4 – A: Consultas EVA

SELECT us.idnumber as Cedula , count(CONCAT(us.idnumber,lo.module,lo.`action`))as
(*<nombre de la actividad>*)

From l as lo, mdl_user as us

Where lo.userid = us.id and lo.module like "*<nombre de la actividad>*"and lo.`action` like
"*<nombre de la acción>*" and us.idnumber in (*<número de cédulas de los estudiantes que se requiere la información>*)

Group by

us.idnumber ,lo.module,lo.`action`

Anexo 4 – B: Consultas gestión Ebooks

Select concat(':',us.cedula,':'),apli.nombre, count(concat(us.cedula,' ',apli.nombre))

from bihistoricoaplicacion as ap ,biusuario as us , bidispositivousuario as dis, biaplicacion as
apli

Where ap.id_dis_fisico = dis.id_dis_fisico and dis.id_usuario = us.id_usuario and

us.cedula in (*<número de cédulas de los estudiantes que se requiere la información>*)

and ap.id_usuario = us.id_usuario and ap.id_aplicacion=apli.id

group by us.cedula, apli.nombre

ANEXO 5: DISTRIBUCIÓN POR CENTROS

	N de
Centro	estudiantes

Alamor	4
Alausí	4
Ambato	70
Azogues	10
Bahía De Caráquez	6
Balsas	1
Calceta	2
Cañar	3
Cariamanga	8
Catacocha	1
Cayambe	10
Chone	8
Coca	15
Cuenca	90
Daule	5
Durán	23
El Chaco	4
Esmeraldas	32
Gualaceo	3
Gualaquiza	4
Guaranda	25
Guayaquil	97
Guayaquil - Centenario	31
Huaquillas	8
Ibarra	54
Joya De Los Sachas	4
La Concordia	6
La Troncal	8
Latacunga	40
Limón Indanza	2
Loja	67
Macará	6
Macas	11
Machachi	9
Machala	70
Madrid	6
Manta	20
Méndez	1
Milagro	24
Nueva Loja (La)	41
Otavalo	10
Pasaje	9
Paute	7
Piñas	2
Portovelo	1
Portoviejo	28

Puyo	22
Quevedo	39
Quinindé	4
Quito	179
Quito-Carcelén	68
Quito-San Rafael	61
Quito-Tumbaco	18
Quito-Turubamba	41
Quito-Villaflora	76
Riobamba	47
Salinas	33
Samborondon	16
San Cristóbal	5
San Gabriel	4
San Miguel De Los Bancos	4
Santa Cruz	6
Santa Isabel	6
Santa Rosa	9
Santo Domingo	47
Saraguro	3
Shushufindi	2
Sucúa	3
Tena	8
Tulcán	17
Vinces	1
Yanzatza	3
Zamora	9
Zaruma	8
Zumba	1
<hr/> Total general	<hr/> 1630

ANEXO 6: APLICACIONES

Anexo 6 – A: Aplicaciones que conforman la categorización de las variables de aplicaciones en la Tableta digital

aplicaciones_co municación	aplicaciones _sociales	aplicaciones_pr oductividad	aplicaciones_ educativas	aplicaciones_I ectorUTPL	aplicaciones_m ediosVideos	aplicaciones_m usicaVideo	aplicaciones _compra	aplicaciones _deportes	aplicaciones_ent retenimiento	aplicaciones_ estiloVida
Chrome	Ask.fm	Adobe Acrobat	Duolingo	UTPL Reader	Flixster	DJStudio 5	MercadoLibre	El Canal del Futbol	Aptoide	Homestyler
Cool Navegador	Badoo	Almacenamiento de medios	En-Es Free Tradutor Offline		Giphy for Messenger	Edjing	Metro Launcher	FotMob	ArtRage	
Correo	Facebook	Diapositivas	Espanol		Google Play Movies	Google Play Music			FxGuru	
Firefox	Google+	Docs To Go	Google Zhuyin Input		VivaVideo	Los 40 Principales - Espana			Google Play Games	
Gmail	Instagram	Drive	Intel Education Study		VLC	Metronome Beats			Netflix	
Hangouts	Lite	Dropbox	Kno Textbooks		You Play Player	Music Player			Television gratis	
Hotmail	Pinterest	Foxit PDF	Papyre Desktop		YouTube	Music Player for Pad/Phone			TubeMate	
LINE	Twitter	Google Drive	SPARKvue			Musica relajante spa			Vaca Lechera	
Messenger		Hojas de calculo	UTPL			Nursery Rhymes PRO			You Player Pro	
Navegador		Lector de PDF de Google				PitchLab Lite				
Opera		Mail Reader for MSN Outlook				Play Tube				
Skype		Notes Mobile				SoundCloud				
Talk		OfficeSuite				Spotify				
Telegram		OffiStar				The Voice				
UC Mini		OneDrive				TuneIn Radio				
Viber		OneNote				Your Voice				
WhatsApp		Outlook								
Yahoo Mail		PDF Reader								
		Polaris Office								
		Prezi								

Puffin Free
 Smart Office 2
 Works Office
 WPS Office
 Wunderlist

aplicaciones_fotografia	aplicaciones_libros	aplicaciones_librosObrosConsulta	aplicaciones_librosReferencias	aplicaciones_multimedia	aplicaciones_musicaAudio	aplicaciones_noticiasRevistas	aplicaciones_salud	aplicaciones_viajes	aplicaciones_video tutorias
Fotos	Diccionario Juridico	Mantano Reader Free	Amazon Kindle	Estudio de peliculas	DJStudio 5	Comercio	uSound Ares V	Earth	BallboardCollaborate
Square InstaPic	DL Reader	WordReference	Google Play Libros	Video Downloader	Edjing	feedly			
SquareQuick	DRAE		Java Programming Tutorials	You Player View	Los 40 Principales - Espana	Google Play Kiosco			
WaterfallPhotoFrames	Ebook Reader		Scribd		Metronome Beats				
YouCam Makeup			UB Reader		Music Player				
YouCam Perfect			Wattpad		Music Player for Pad/Phone				
					Musica relajante spa				
					Nursery Rhymes PRO				
					PitchLab Lite				
					Play Tube				
					SoundCloud				
					Spotify				
					The Voice				
					TuneIn Radio				
					Your Voice				

Anexo 6 – B: Juegos que conforman la categorización de las variables de juegos en la Tableta digital

juegos_accion	juegos_aventura	juegos_carreas	juego_cartas	juegos_clasicos	juegos_creatividad	juegos_deportes	juegos_educativos	juegos_estrategia	juegos_musica	juegos_ocasionales
Agar.io	Angry Anaconda Attack 3D	Asphalt 8	O&C Duels	Hero Racing Alliance	Monster Hair	8BallPool	Balloon Pop	Age of Civilizations Lite	Music Hero	AbPop
Banana Kong	Angry Birds	Bike Race	Solitario	SG: DeadZone		Bowling King	Disney Junior Play	Boom Beach		Burger Shop FREE
Bus Rush	Angry Birds 2	Buscador RAE	UNOfriends	Sky Force		Dream League Soccer	Doc Mc Stuffins Lamb	Clash of Clans		Candy Crush Saga
Dark Waters	DragonCity	Car Traffic Race		Snake King		Fifa15:UT	English for Kids	Clash of Lords 2: Espanol		Candy Crush Soda
DH 2014	EggCar	Cars				FTS15	Learn To Write	Contra Zombies		Free Fruit Cut
Dino Hunter	Epic	CarX Drift Racing				Tennis 3D	My Boo	DomiNations		Hay Day
Dragon ML	Family Guy	Earn to Die 2				Top Eleven	Papermaking with Kiki	Invasion		Ice Princess
Dumb Ways 2	Flappy Bird	Extreme Car Driving Simulator				UFC		Plants Vs Zombies 2		Mi Talking Angela
Gunners Battle Desert Storm	Ice Age Adventures	GoGoThomas						Plants vs. Zombies Free		Mi Talking Tom
Jurassic Nights	Lep's World	Hill Climb Racing								Pou
Jurassic Park Builder	Lep's World 2	Horizon Chase								Princess Nail Salon
Just Shout	Looney Tunes A Corer	Moto Traffic Race								Springfield
Killer Bean Unleashed	Masha y el oso	Offroad Legends 2								SZ3
Lose Weight	Minion Rush	Racing Fever								
Marvel	Minions	Rally Racer Dirt								
Mutants	Peludos y Tropezosos	Real Off-Road 4x4								
Pac-Man Dash	Sky Whale	Traffic								
PPSSPP	Super Vito Adventure	Traffic Racer								
Rayman Jungle Run		Trial Xtreme 3								
SoulCraft		Turbo Car Racing								
Spider-Man		Turbo SE								
Stickman Battlefields		Zombie Roadkill								

Temple Run
 Temple Run 2
 Zombie Squad
 Zombie Crush
 Zombie Reaper

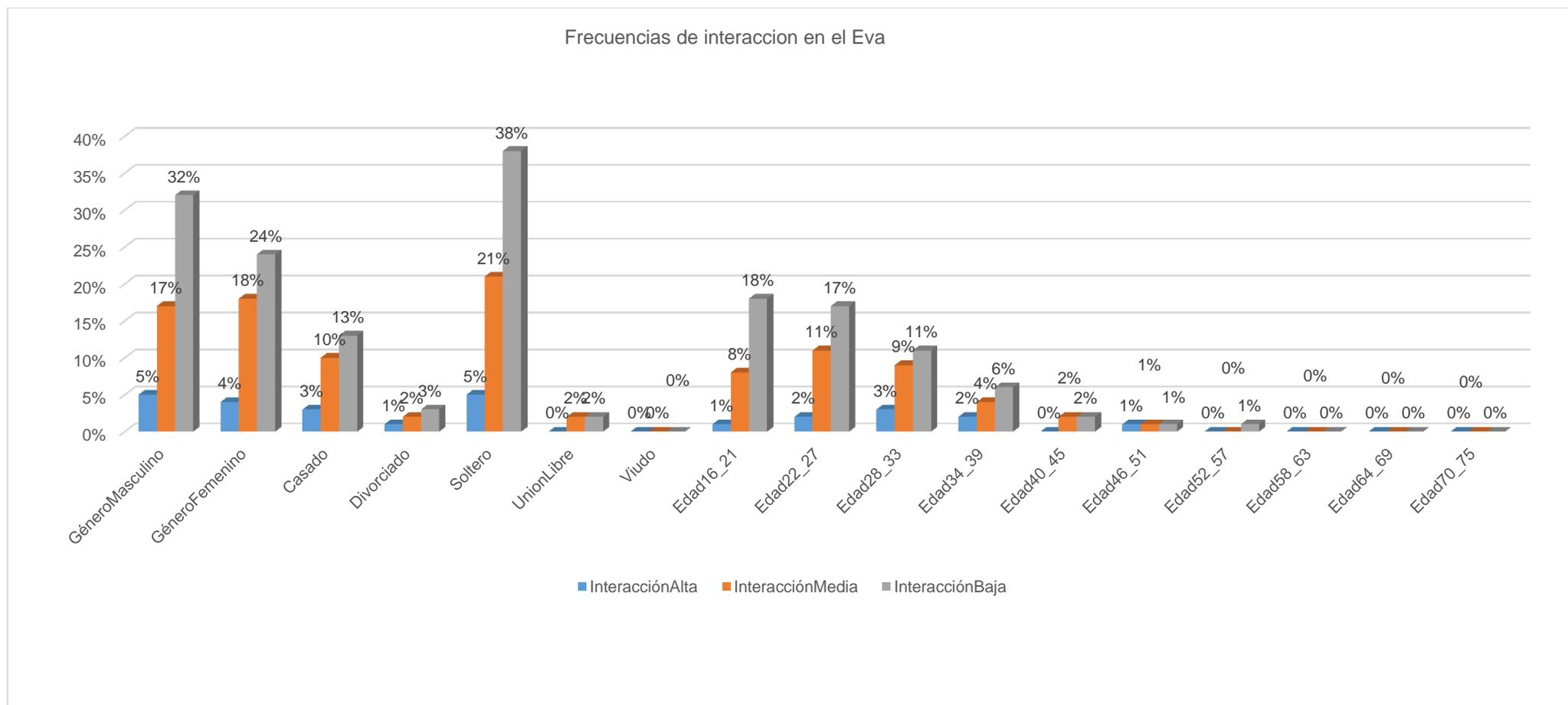
juegos_preguntasRespuestas juegos_puzzles juegos_rompecabeza juegos_salaJuegos juegos_simulacion

Cierto o falso	Apensar	Cachorros - Rompecabezas	Fruit Ninja Free	Airport Parking
Preguntados	Bejeweled	Charlie Hop	Geometry Dash	Bus Simulator 2015
	Diamond Digger Saga	Frozen Free Fall	Geometry Dash Lite	El resurgir de Mema
	Farm Heroes Saga	FROZEN QUEENS	Glow Hockey	FreePlay
	Marble Legend	Logic Dots	Hungry Shark	Ice Age Village
	Pet Rescue Saga	Sopa de Letras	Piano Tiles 2	Ice Cream
			Pimple POP	MOM S2
			Prize Claw	My Burger Shop 2
			Sonic Dash	Paradise Island 2
			Stupid Zombies	Police Bus Prisoner Transport
			Subway Surf	Public Transport Simulator
			Super Chaves World	Sport Car Simulator
			WaveRun	Wild Wolf Attack Simulator 3D
			ZigZag	
			Zombie Diary	
			Zombie Smasher	
			Zombie Tsunami	

aplicaciones_comunicación	aplicaciones_sociales	aplicaciones_productividad	aplicaciones_educativas	aplicaciones_lector UTPL
Chrome	Ask.fm	Adobe Acrobat	Duolingo	UTPL Reader
Cool Navegador	Badoo	Almacenamiento de medios	En-Es Free Tradutor Offline	
Correo	Facebook	Diapositivas	Espanol	
Firefox	Google+	Docs To Go	Google Zhuyin Input	
Gmail	Instagram	Drive	Intel Education Study	
Hangouts	Lite	Dropbox	Kno Textbooks	
Hotmail	Pinterest	Foxit PDF	Papyre Desktop	
LINE	Twitter	Google Drive	SPARKvue	
Messenger		Hojas de calculo	UTPL	
Navegador		Lector de PDF de Google		
Opera		Mail Reader for MSN		
Skype		Outlook		
Talk		Notes Mobile		
Telegram		OfficeSuite		
UC Mini		OffiStar		
Viber		OneDrive		
WhatsApp		OneNote		
Yahoo Mail		Outlook		
		PDF Reader		
		Polaris Office		
		Prezi		
		Puffin Free		
		Smart Office 2		
		Works Office		
		WPS Office		
		Wunderlist		

ANEXO 7: FRECUENCIAS DE INTERACCIÓN

Anexo 7 – A: Género, estado_civil, edad EVA



Anexo 7 – B: Género, estado_civil, edad Tableta digital

