



UNIVERSIDAD TÉCNICA PARTICULAR DE LOJA

*La Universidad Católica de Loja*

AREA TÉCNICA

TÍTULO DE INGENIERO EN INFORMÁTICA

**Análisis de los usos de la tecnología en los estudiantes de la  
Universidad PUCE - Sede Ibarra.**

TRABAJO DE TITULACIÓN

AUTOR: Torres Carrión, Jaime Fabián

DIRECTOR: Guamán Bastidas, Franco Olivio, Mgtr.

CENTRO UNIVERSITARIO QUITO

2016

## **APROBACIÓN DEL DIRECTOR DEL TRABAJO DE TITULACIÓN**

Magíster.

Franco Olivio Guamán Bastidas

### **DOCENTE DE TITULACION**

De mi consideración:

El presente trabajo de titulación: Análisis de los usos de la tecnología en los estudiantes de la Universidad PUCE - Sede Ibarra, realizado por Jaime Fabián Torres Carrión ha sido orientado y revisado durante su ejecución, por cuanto se aprueba la presentación del mismo.

Loja, Abril del 2016

f) .....

## DECLARACIÓN DE AUTORÍA Y CESIÓN DE DERECHOS

Yo Jaime Fabián Torres Carrión declaro ser autor del presente trabajo de titulación: Análisis de los usos de la tecnología en los estudiantes de la Universidad PUCE - Sede Ibarra, siendo el director de titulación asignado el Ingeniero Franco Olivio Guamán Bastidas; y eximo expresamente a la Universidad Técnica Particular de Loja y a sus representantes legales de posibles reclamos y acciones legales. Además certifico que las ideas, conceptos, procedimientos y resultados vertidos en el presente trabajo investigativo, son de mi exclusiva responsabilidad.

Adicional 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) .....

Autor: Torres Carrión, Jaime Fabián

Cédula: 1104587124

## DEDICATORIA

A Dios y a la Virgen por ayudarme siempre; a mis padres Jaime y Elsa que han estado conmigo en todo el camino recorrido y me han apoyado siempre sin cuestionarme, que han sembrado en mi todos los valores que necesito como persona para ser feliz y hacer feliz a quienes me rodean; a mi esposa Mariela y a mi hija Marielita que con su amor llenan de calor nuestro hogar, y que me han permitido sacrificar el tiempo que tengo que pasar junto a ellas, para dedicarlo a mis estudios; a mi hermano Pablo que con su ejemplo y ganas de superación constante, me ha encaminado y me ha dado ánimo para seguir adelante, y es a quien debo muchos de mis logros; a mis demás hermanos Tania, Sonia, Daniel, Elizabeth, Verónica y Ángel que siempre están pendientes de mí, y que con sus sabios consejos me han sabido enseñar muchas cosas valiosas a lo largo de mi vida.

## **AGRADECIMIENTO**

A todos mis profesores que a lo largo de mi carrera me han sabido enseñar cada uno de los temas que me han servido para desempeñarme en mi trabajo y para poder realizar este trabajo de investigación, en especial a la Dra. Maria Jose Rubio que me apoyo permitiéndome levantar la información en la PUCE-SI, al Mgtr. Juan Figueroa que me ha sabido impulsar y dar ánimo para finalizar mi carrera, al Mgtr. Juan Carlos Torres y al Mgtr. Franco Guamán que me han guiado en el transcurso de este trabajo de investigación.

Un agradecimiento especial a todos mis compañeros de trabajo, que me han brindado su confianza siempre y me han sabido apoyar e impulsar en mi carrera laboral, en especial al Ingeniero Guillermo Moreano, al Ingeniero Jeffry Arroba, a la Ingeniera Marcela Bermudez y muchos otros más con los que he tenido el gusto de compartir más de ocho horas diarias por muchos años.

## INDICE DE CONTENIDOS

---

APROBACIÓN DEL DIRECTOR DEL TRABAJO DE TITULACIÓN.....	II
DECLARACIÓN DE AUTORÍA Y CESIÓN DE DERECHOS.....	III
DEDICATORIA .....	IV
AGRADECIMIENTO .....	V
RESUMEN.....	XXI
ABSTRACT .....	XXII
CAPITULO I - INTRODUCCIÓN .....	1
1. Introducción .....	2
1.1. Estructura del Documento .....	2
1.2. Visión y alcance.....	3
1.3. Objetivos .....	5
1.4. Preguntas de investigación e hipótesis respectivas .....	5
CAPÍTULO II - ESTADO DEL ARTE.....	7
2. Estado del Arte .....	8
2.1. Realidad Tecnológica actual en el Ecuador .....	8
2.2. Posibilidades de acceso a la educación superior.....	13
2.3. Evolución universitaria desde una perspectiva Tecnológica. ....	14
2.4. La Web y el impacto que ha ocasionado en la Educación Superior. ....	17
2.5. La brecha digital en la sociedad de la información.....	20

2.6.	El rendimiento o éxito académico frente al uso de internet. ....	22
2.7.	El nivel socioeconómico frente al uso de internet. ....	24
CAPÍTULO III - MARCO TEORICO .....		27
3.	Marco Teórico.....	28
3.1.	La brecha digital .....	28
3.2.	Factores que influyen en la brecha digital.....	29
3.2.1.	Los ingresos económicos. ....	29
3.2.2.	La edad. ....	30
3.2.3.	El Género. ....	31
3.2.4.	El idioma.....	32
3.2.5.	La geografía. ....	32
3.2.6.	La educación. ....	33
3.2.7.	La integridad física.....	33
3.3.	Influencia del internet en el rendimiento académico universitario.....	35
3.4.	Tecnologías como herramienta académica en la Universidad. ....	36
3.5.	Aprendizaje Digital.....	39
CAPÍTULO IV - MINERÍA DE DATOS.....		42
4.	Minería de Datos .....	43
4.1.	Áreas o Mercados en lo que se aplica la minería de datos .....	43
4.1.1.	Banca y Comercio .....	44
4.1.2.	Medicina y Farmacia.....	44
4.1.3.	Seguridad y Detección de Fraude.....	44

4.1.4.	Sistema Educativo .....	44
4.1.5.	Geología, Minería y Agricultura.....	44
4.1.6.	Ciencias Ambientales .....	44
4.1.7.	Ciencias Sociales .....	45
4.2.	Tipos de Bases de Datos.....	45
4.3.	Modelos y Tareas de Minería de Datos .....	46
4.3.1.	Modelos Descriptivos y sus Tareas .....	46
4.3.2.	Modelos Predictivos y sus Tareas .....	47
4.4.	Técnicas o Métodos de Minería de Datos.....	47
4.4.1.	Métodos Bayesianos .....	48
4.4.2.	Reglas de Asociación y Dependencia.....	48
4.4.3.	Métodos relacionales y estructurales.....	48
4.4.4.	Métodos basados en casos y en vecindad. ....	49
4.4.5.	Árboles de decisión. ....	49
4.4.6.	Redes Neuronales.....	49
4.4.7.	Análisis Factorial (factorización). ....	50
4.4.8.	Modelización estadística paramétrica. ....	50
4.4.9.	Modelización estadística no paramétrica. ....	51
4.5.	Algoritmos de Minería de Datos.....	51
4.5.1.	Algoritmos evolutivos y reglas difusas. ....	51
4.5.2.	Algoritmos de minería de datos en función del objetivo .....	52
4.5.3.	Algoritmos de clustering o agrupamiento.....	52

4.5.4.	Algoritmos de clasificación.....	57
4.5.5.	Algoritmos de asociación.....	58
4.5.6.	Otros Algoritmos.....	58
4.6.	Herramientas de Minería de Datos .....	59
4.6.1.	Weka .....	59
4.6.2.	Spss Clementine .....	59
4.6.3.	ODMS (Oracle Data Mining suite).....	60
4.6.4.	Kepler .....	60
4.6.5.	RapidMiner .....	60
4.6.6.	Db2 Intelligent Miner.....	60
4.6.7.	Statistica Data Miner.....	61
4.6.8.	Orange .....	61
4.7.	Procesos para la aplicación de la Minería de Datos.....	61
4.7.1.	Proceso KDD.....	61
4.7.2.	Proceso SEMMA .....	63
4.7.3.	Proceso CRISP-DM.....	64
CAPÍTULO V - MARCO METODOLÓGICO .....		67
5.	Marco Metodológico .....	68
5.1.	Tipo y Nivel.....	68
5.2.	Método y Diseño.....	68
5.3.	Población y Muestra. ....	69
5.4.	Técnicas e Instrumentos de recolección de datos.....	69

5.5.	Análisis de datos .....	70
5.5.1.	Análisis Factorial .....	71
5.5.2.	Análisis Cluster .....	71
5.5.3.	Análisis de regresión Logística .....	72
5.6.	Definición de variables a utilizar.....	72
5.6.1.	Variable de nivel de ingreso.....	72
5.6.2.	Variables de Usos de Internet en actividades académicas. ....	73
5.6.3.	Variables de Usos de Internet en actividades de entretenimiento. ....	73
5.6.4.	Variables de rendimiento académico .....	73
5.7.	Herramienta de minería de datos a utilizar .....	74
CAPÍTULO VI - DESARROLLO .....		77
6.	Desarrollo .....	78
6.1.	Fase I - Compresión del Negocio.....	78
6.1.1.	Determinación de objetivos del negocio.....	78
6.1.2.	Valoración de la situación .....	79
6.1.3.	Determinación de objetivos de Minería de Datos.....	81
6.1.4.	Creación del plan del proyecto .....	81
6.2.	Fase II - Compresión de los datos .....	82
6.2.1.	Recolección de los datos iniciales .....	82
6.2.2.	Descripción de los datos.....	82
6.2.3.	Exploración de datos. ....	89
6.2.4.	Verificación de calidad de los datos.....	124

6.3.	Fase III – Preparación de los datos .....	125
6.3.1.	Selección de Datos.....	125
6.3.2.	Limpieza de los Datos .....	126
6.3.3.	Estructuración e Integración de los Datos.....	126
6.3.4.	Formateo de los Datos .....	128
6.4.	Fase IV – Modelado.....	128
6.4.1.	Selección de la técnica de modelado.....	128
6.4.2.	Construcción del modelo .....	129
6.5.	Fase V – Evaluación.....	153
CAPÍTULO VII - DISCUSIÓN Y CONCLUSIONES .....		155
7.	Discusión y Conclusiones .....	156
7.1.	Incidencia del nivel de Ingreso frente al uso de internet en actividades académicas.....	156
7.2.	Incidencia del nivel de Ingreso frente al uso de internet en actividades de entretenimiento. ....	158
7.3.	Incidencia del uso de internet en actividades académicas frente al rendimiento académico. ....	159
7.4.	Incidencia del uso de internet en actividades de entretenimiento frente al rendimiento académico. ....	160
7.5.	Conclusiones.....	161
RECOMENDACIONES.....		163
8.	Recomendaciones.....	164
BIBLIOGRAFIA.....		165

ANEXOS.....	174
-------------	-----

## INDICE DE FIGURAS

---

Figura 1. The NRI Ranking FEM 2012 (142 países evaluados). .....	8
Figura 2. Personas que utilizan computadora por grupos de edad en Ecuador.....	11
Figura 3. Acceso a Internet según el Área en Ecuador. ....	12
Figura 4. Porcentaje de personas Analfabetas Digitales en Ecuador.....	13
Figura 5. Número de publicaciones en las Universidades de Ecuador hasta el 2012..	15
Figura 6. Número de publicaciones en SCOPUS Universidades de Ecuador. ....	16
Figura 7. Número de publicaciones en SCOPUS Universidades de Ecuador en el 2015. .....	16
Figura 8. La evolución de las TIC en la educación. ....	18
Figura 9. Hogares con acceso a Internet según quintil de ingresos. ....	25
Figura 10A. Dos enfoques de docentes Universitarios(Antes). ....	41
Figura 11. Error cuadrático K-means. ....	54
Figura 12. Ciclo de Vida del proceso CRISP-DM. ....	65
Figura 13. Grafica de distribución Ingresos Económicos.....	90
Figura 14. Grafica de distribución de ¿Cuántas veces a la semana ingresa a la plataforma virtual de su universidad?.....	92
Figura 15. Grafica de distribución de ¿Cuántas consultas a través de TIC's les hace a sus profesores cada mes?.....	94
Figura 16. Grafica de distribución de ¿Cuántas consultas a través de TIC's realiza a sus compañeros por mes?.....	96

Figura 17. Grafica de distribución de ¿Cuántos recursos educativos descarga de la plataforma virtual cada mes? .....	98
Figura 18. Grafica de distribución de ¿Cuántos videos académicos mira en youtube cada mes? .....	100
Figura 19. Grafica de distribución de ¿En cuántos foros virtuales participa cada mes? .....	102
Figura 20. Grafica de distribución de ¿Cuántos post o tweets sobre temas académicos realiza en las redes sociales por mes? .....	104
Figura 21. Grafica de distribución de ¿Cuántas horas chatea sobre temas académicos cada mes? .....	106
Figura 22. Grafica de distribución de ¿Cuántas horas busca información académica en internet cada mes? .....	108
Figura 23. Grafica de distribución de ¿Cuántas horas utiliza la biblioteca virtual de la universidad cada mes? .....	110
Figura 24. Grafica de distribución de ¿Cuántas horas a la semana chatea por diversión? .....	112
Figura 25. Grafica de distribución de ¿Cuántas horas a la semana utiliza redes sociales?.....	114
Figura 26. Grafica de distribución de ¿Cuántas horas a la semana utiliza juegos en línea?.....	116
Figura 27. Grafica de distribución de ¿Cuántas horas a la semana descarga música, videos y programas?.....	118
Figura 28. Grafica de distribución de ¿Cuántos videos para entretenimiento mira en youtube cada semana?.....	120
Figura 29. Grafica de distribución de ¿En cuántas asignaturas se matriculó? .....	122
Figura 30. Grafica de distribución de ¿Cuántas asignaturas aprobó? .....	124

Figura 31. Operaciones estadísticas como resultado del nuevo atributo.....	127
Figura 32. Resultado aplicación de Tabla de decisión a DataSet "Uso de internet en actividades académicas".....	132
Figura 33. Resultado aplicación de Tabla de decisión a DataSet "Uso del internet en actividades de entretenimiento" .....	132
Figura 34. Resultado aplicación de algoritmo EM a DataSet "uso del internet en actividades académicas".....	133
Figura 35. Resultado aplicación de algoritmo EM a DataSet "uso del internet en actividades de entretenimiento" .....	134
Figura 36. Resultado algoritmo <i>SimpleKmeans</i> a DataSet "uso del internet en actividades académicas".....	136
Figura 37. Resultado algoritmo <i>SimpleKmeans</i> a DataSet "uso del internet en actividades de entretenimiento" .....	138
Figura 38. Coeficientes del modelo de regresión logística multinomial (ingreso económico y uso de Internet en actividades académicas) .....	141
Figura 39. Resultado aplicación de técnica de regresión logística multinomial (ingreso económico y uso de Internet en actividades académicas) .....	142
Figura 40. Coeficientes del modelo de regresión logística multinomial (ingreso económico y uso de Internet en actividades de entretenimiento) .....	144
Figura 41. Resultado aplicación de técnica de regresión logística multinomial (ingreso económico y uso de Internet en actividades de entretenimiento) .....	144
Figura 42. Coeficientes del modelo de regresión logística multinomial (uso de Internet en actividades académicas y rendimiento académico).....	146
Figura 43. Resultado aplicación de técnica de regresión logística multinomial (uso de Internet en actividades académicas y rendimiento académico).....	147
Figura 44. Coeficientes del modelo de regresión logística multinomial (uso de Internet en actividades de entretenimiento y rendimiento académico) .....	149

Figura 45. Resultado aplicación de técnica de regresión logística multinomial (uso de Internet en actividades de entretenimiento y rendimiento académico) .....	150
Figura 46. Grafica de relación de ingreso económico - uso de Internet en actividades académicas .....	151
Figura 47. Grafica de relación de ingreso económico - uso de Internet en actividades de entretenimiento .....	151
Figura 48. Grafica de relación del uso de Internet en actividades académicas - rendimiento académico.....	152
Figura 49. Grafica de relación del uso de Internet en actividades de entretenimiento - rendimiento académico.....	153

## INDICE DE TABLAS

---

Tabla 1. The NRI Ranking 2008 - 2014 y Proyección al 2016.....	9
Tabla 2. Diseño aplicado al trabajo de investigación.....	68
Tabla 3. Cuadro comparativo de algoritmos en herramientas de Minería de Datos ....	76
Tabla 4. Plan de Ejecución del Proyecto.....	81
Tabla 5. Información captada mediante encuestas a estudiantes Universitarios .....	83
Tabla 6. Uso internet en actividades académicos - DataSet .....	88
Tabla 7. Uso internet en actividades de entretenimiento – DataSet .....	88
Tabla 8. Rendimiento académico – DataSet.....	89
Tabla 9. Frecuencias de ingresos económicos .....	90
Tabla 10. Operaciones Estadísticas de ingresos económicos.....	90
Tabla 11. Frecuencias de ¿Cuántas veces a la semana ingresa a la plataforma virtual de su universidad?.....	91
Tabla 12. Operaciones Estadísticas ¿Cuántas veces a la semana ingresa a la plataforma virtual de su universidad?.....	91
Tabla 13. Frecuencias de ¿Cuántas consultas a través de TIC's les hace a sus profesores cada mes? .....	93
Tabla 14. Operaciones Estadísticas ¿Cuántas consultas a través de TIC's les hace a sus profesores cada mes?.....	94
Tabla 15. Frecuencias de ¿Cuántas consultas a través de TIC's realiza a sus compañeros por mes? .....	95
Tabla 16. Operaciones Estadísticas ¿Cuántas consultas a través de TIC's realiza a sus compañeros por mes? .....	96

Tabla 17. Frecuencias de ¿Cuántos recursos educativos descarga de la plataforma virtual cada mes?.....	97
Tabla 18. Operaciones Estadísticas ¿Cuántos recursos educativos descarga de la plataforma virtual cada mes? .....	98
Tabla 19. Frecuencias de ¿Cuántos videos académicos mira en youtube cada mes? 99	
Tabla 20. Operaciones Estadísticas ¿Cuántos videos académicos mira en youtube cada mes? .....	99
Tabla 21. Frecuencias de ¿En cuántos foros virtuales participa cada mes? .....	101
Tabla 22. Operaciones Estadísticas ¿En cuántos foros virtuales participa cada mes? .....	101
Tabla 23. Frecuencias de ¿Cuántos post o tweets sobre temas académicos realiza en las redes sociales por mes?.....	103
Tabla 24. Operaciones Estadísticas ¿Cuántos post o tweets sobre temas académicos realiza en las redes sociales por mes? .....	103
Tabla 25. Frecuencias de ¿Cuántas horas chatea sobre temas académicos cada mes? .....	105
Tabla 26. Operaciones Estadísticas ¿Cuántas horas chatea sobre temas académicos cada mes? .....	106
Tabla 27. Frecuencias de ¿Cuántas horas busca información académica en internet cada mes? .....	107
Tabla 28. Operaciones Estadísticas ¿Cuántas horas busca información académica en internet cada mes? .....	108
Tabla 29. Frecuencias de ¿Cuántas horas utiliza la biblioteca virtual de la universidad cada mes? .....	109
Tabla 30. Operaciones Estadísticas ¿Cuántas horas utiliza la biblioteca virtual de la universidad cada mes? .....	110

Tabla 31. Frecuencias de ¿Cuántas horas a la semana chatea por diversión?.....	111
Tabla 32. Operaciones Estadísticas ¿Cuántas horas a la semana chatea por diversión? .....	112
Tabla 33. Frecuencias de ¿Cuántas horas a la semana utiliza redes sociales?.....	113
Tabla 34. Operaciones Estadísticas ¿Cuántas horas a la semana utiliza redes sociales?.....	114
Tabla 35. Frecuencias de ¿Cuántas horas a la semana utiliza juegos en línea? .....	115
Tabla 36. Operaciones Estadísticas ¿Cuántas horas a la semana utiliza juegos en línea?.....	116
Tabla 37. Frecuencias de ¿Cuántas horas a la semana descarga música, videos y programas? .....	117
Tabla 38. Operaciones Estadísticas ¿Cuántas horas a la semana descarga música, videos y programas?.....	118
Tabla 39. Frecuencias de ¿Cuántos videos para entretenimiento mira en youtube cada semana?.....	119
Tabla 40. Operaciones Estadísticas ¿Cuántos videos para entretenimiento mira en youtube cada semana?.....	120
Tabla 41. Frecuencias de ¿En cuántas asignaturas se matriculó?.....	121
Tabla 42. Operaciones Estadísticas ¿En cuántas asignaturas se matriculó?.....	121
Tabla 43. Frecuencias de ¿Cuántas asignaturas aprobó? .....	123
Tabla 44. Operaciones Estadísticas ¿Cuántas asignaturas aprobó? .....	123
Tabla 45. Técnicas y Algoritmos DM.....	129
Tabla 46. Centroides de agrupación de atributos de DataSet Actividades Académicas .....	135

Tabla 47. Centroides de agrupación de atributos de DataSet Actividades de entretenimiento.....	137
Tabla 48. Matriz de confusión ingreso económico y uso del internet en actividades académicas .....	142
Tabla 49. Matriz de confusión (ingreso económico y uso del internet en actividades de entretenimiento).....	145
Tabla 50. Matriz de confusión (uso del internet en actividades académicas y rendimiento académico).....	147
Tabla 51. Matriz de confusión (uso del internet en actividades de entretenimiento y rendimiento académico).....	150

## RESUMEN

En Ecuador como en todos los países, es cada vez más visible la evolución tecnológica que estamos viviendo, esto se lo puede apreciar de manera clara en la educación. La tecnología en la educación puede afectar positiva o negativamente de acuerdo a lo factores que incidan en el uso de la misma. En la educación superior que es el foco de este trabajo, se ha visto que factores como los ingresos económicos influyen en la manera que los estudiantes utilizan la tecnología y a su vez la incidencia que este uso tiene frente a su éxito académico. Para confirmar estos factores se procedió a levantar información mediante encuestas a los estudiantes de la Pontificia Universidad Católica del Ecuador – Sede Ibarra, luego mediante la metodología CRISP-DM, y la herramienta WEKA se logró crear modelos de minería de datos que nos permitieron conocer la incidencia que tiene el ingreso económico del estudiante frente al uso del internet en actividades académicas y en actividades de entretenimiento; así como también la incidencia que tienen estos usos de internet frente al éxito académico del estudiante.

**Palabras Claves:** Educación Superior, tecnología, minería de datos, CRISP-DM, internet, éxito académico, ingresos económicos, PUCE-SI.

## ABSTRACT

In Ecuador as in all countries, is increasingly visible technological evolution we are living, this is what can be seen clearly in education. Technology in education can affect positively or negatively according to the factors influencing the use of it. In higher education that is the focus of this work, it has been found that factors such as economic income influence how students use technology and in turn the impact that this application has face their academic success. To confirm these factors proceeded to gather information through surveys of students at the Pontificia Universidad Católica del Ecuador – Ibarra headquarters, then by the CRISP-DM methodology, and the WEKA tool is able to create models of data mining that allowed us to know the incidence of the economic income of the student over the use of internet in academic activities and entertainment; as well as incidence that these uses internet against student academic success.

**Keywords:** Higher education, technology, data mining, CRISP-DM, internet, academic success, economic income, PUCE-SI.

## **CAPITULO I - INTRODUCCIÒN**

## **1. Introducción**

### **1.1. Estructura del Documento**

El presente documento se divide en siete capítulos, en los mismos que se presenta de forma organizada el trabajo de experimentación sobre el uso de la tecnología de información y comunicación en la Educación Superior, particularmente en la Pontificia Universidad Católica del Ecuador Sede Ibarra (PUCE-SI) que es la Institución Educativa en donde se aplica esta investigación.

En el Capítulo I se presenta la visión y alcance de este trabajo de investigación, los objetivos a alcanzar en el mismo, y las preguntas de investigación con sus respectivas hipótesis a discutir y resolver en esta investigación, apoyados en modelos y técnicas que nos brinda la minería de datos aplicando la metodología CRISP-DM (Cross-Industry Standard Process for Data Mining).

El Capítulo II comprende el estado de la cuestión, sobre la situación tecnológica en el Ecuador en relación al uso del Internet, enfocándonos en la población estudiantil de nivel superior. El minucioso trabajo ha permitido establecer los alcances en cuanto al uso de tecnologías de minería de datos para este tipo de estudios, las metodologías, modelos y técnicas más utilizadas; y que se alinean de mejor manera en nuestro trabajo de investigación. También nos ha ayudado a identificar por medio de investigaciones emprendidas por otros estudiosos, la importancia que actualmente tiene el internet en nuestro entorno social y los resultados positivos o negativos que podemos obtener de esta tecnología de acuerdo al uso que le demos a la misma.

El Capítulo III corresponde al marco teórico que de forma estratégica se ha dividido en dos apartados. En este capítulo que comprende la primera parte del marco teórico, se expone los principales factores que influyen para la presencia de la brecha digital; la influencia que tiene el internet en el rendimiento académico de los estudiantes universitarios; las tecnologías cuyas bondades se alinean para ser usadas como herramientas académicas dentro y fuera del aula de clase; y el aprendizaje digital tomado como un método de estudio habitual en la mayoría de centros de educación superior.

En el Capítulo IV se complementa el marco teórico, detallando en este apartado todo lo referente a minería de datos, en el que se contempla los modelos, tareas, técnicas,

algoritmos, reconocimiento de patrones, clustering, factorización y regresiones aplicando minería de datos y la metodología a utilizar para aplicarlos. También se encuentran detallados las herramientas de minería de datos más utilizadas, y las áreas o mercados en donde más se aplican las mismas.

El Capítulo V, corresponde al marco metodológico, en el cual se describe información de la metodología aplicada en el desarrollo de la investigación. La población de estudio corresponde al alumnado de la Pontificia Universidad Católica del Ecuador Seda Ibarra (PUCE-SI), donde se obtuvo una muestra representativa de quinientos estudiantes de las diferentes carreras académicas que oferta este centro de educación superior; se detalla así mismo el procedimiento experimental y metodológico seguido durante todas las etapas de trabajo de campo y de posterior análisis, donde destacan las técnicas de minería de datos y herramientas como Weka para el tratamiento de la información.

En el Capítulo VI, se encuentra la ejecución o desarrollo de la investigación aplicando la metodología de minería de datos CRISP-DM (Cross-Industry Standard Process for Data Mining), utilizando la herramienta Weka para el modelado de la data levantada mediante las encuestas aplicadas a una muestra de 500 estudiantes de las diferentes carreras de la PUCE-SI.

El Capítulo VII, resume la discusión de resultados y las conclusiones obtenidas en este trabajo de investigación.

## **1.2. Visión y alcance**

En el Ecuador como en todos los países del mundo es cada vez más visible la evolución tecnológica que estamos viviendo, así como los cambios sociales que colateralmente se están dando producto de esta evolución. Una de las herramientas tecnológicas que ha pasado de ser un lujo a ser una necesidad para el conocimiento es el Internet, el cual según (Real Academia Española, 2001) ha sido definido como *“Red informática mundial, descentralizada, formada por la conexión directa entre computadoras mediante un protocolo especial de comunicación”*.

La evolución de los sistemas de comunicación a partir de la invención del telégrafo hasta la actualidad, preparó el escenario para la llegada del internet que ha evolucionado el mundo de la informática y las comunicaciones (Leiner et al., 2009).

La forma más frecuente de conexión a internet es a través de los sitios Web; los cuales hasta hace pocos años atrás estaban compuesto de solo texto (Zhang, Zhang, Yang, Cheng, & Zhou, 2008). Pero la evolución de la tecnología ha permitido evolucionar la forma de programar y administrar los sitios web, pasando de páginas estáticas y sin interacción directa con el usuario a páginas dinámicas que administran mucha información de los usuarios, obligando a la consideración de nuevas aristas necesarias para su evolución como son: Seguridad de la información, servicios de hosting, redes de telecomunicación, entre otros.

La evolución del internet le permitió ir ganando terreno en los sectores sociales a nivel mundial; uno de los sectores en donde el uso de esta tecnología ha generado un gran impacto social y económico es el educativo superior y es donde se enfoca esta investigación. En la actualidad los estudiantes universitarios y sus docentes tienen un importante nivel de uso del internet y el involucramiento de los mismos con dicho uso desprende algunas aristas que es necesario analizar pausadamente para comprender integralmente los afectos sociales y económicos que impactarían directa o indirectamente; dichas aristas contemplarían: nivel de manejo de tecnologías por parte de los docentes, las brechas socio-económicas existentes entre los estudiantes, limitante de acceso a internet y aprovechamiento correcto o incorrecto de dicho acceso por parte de estudiantes y docentes (I. Guerrero & Kalman, 2010). Si bien nuestro país ha emprendido nuevas metodologías de control y administración de las Universidades a nivel nacional, aun no se cuenta con indicadores de evaluación que permita valorar el porcentaje de aprovechamiento de las tecnologías por los estudiantes de nivel superior; limitando con ello a visualizar a nivel país, necesidades latentes para el mejoramiento de la educación del siglo XXI, en donde el uso de las tecnologías se da desde muy temprana edad y es en donde hay que culturizar y fomentar el aprovechamiento positivo de las mismas para que se vayan explotando en niveles progresivos en todo el plan de estudios primarios, secundarios, superiores y post-profesionales.

Esta investigación tiene como resultado esperado, lograr por medio de variables aplicables conocer la relación existente entre el rendimiento académico alcanzado por los estudiantes de nivel superior, frente al uso que le dan los mismos a las tecnologías actuales. Para ello se ha realizado un análisis de relación de variables tomando en cuenta factores importantes como, nivel de ingresos, edad, género, nivel de conocimiento sobre tecnologías, frecuencia de uso e interacción con las diferentes

herramientas tecnológicas actuales tanto estudiantes como docentes, y el rendimiento académico alcanzado.

En línea al párrafo anterior, uno de los factores que influye significativamente a la brecha digital entre los estudiantes es el nivel de ingresos, mismo que va a ser foco de nuestro estudio, pues por la incapacidad de asumir los costes que conllevan tener acceso a la tecnología muchos estudiantes carecen de este acceso, lo que provoca que se sientan limitados o en desventaja frente a los otros estudiantes que tienen acceso a tecnologías por tener un mejor nivel de ingresos familiar.(Cusí, 2003)

La materia prima para este trabajo (los datos) fue levantada mediante la aplicación de encuestas a una muestra de 500 estudiantes de la Universidad Pontificia Católica del Ecuador Sede Ibarra (PUCE-SI) de las diferentes carreras que esta prestigiosa Institución oferta. Y los mismos serán estudiados con la finalidad de lograr el resultado esperado, mismo que fue descrito en los párrafos anteriores.

### **1.3. Objetivos**

Los objetivos trazados para la presente investigación son:

- Analizar e identificar segmentos de usuarios de acuerdo al uso que le den los mismos a las tecnologías para sus fines académicos.
- Establecer el nivel de incidencia de las variables socio-económicas frente al uso del internet a través de modelos estadísticos.
- Determinar el nivel de impacto que presenta el uso de las tecnologías que aplican cada segmento de estudiantes frente al rendimiento académico alcanzado por los mismos.
- Aplicar minería de datos para la demostración de la correcta aplicabilidad de los métodos estadísticos en el trabajo de investigación.

### **1.4. Preguntas de investigación e hipótesis respectivas**

Las preguntas estipuladas para la presente investigación son:

- a. ¿Incide el nivel de ingreso frente a la forma en cómo usen el internet los estudiantes de nivel superior?

### Hipótesis

- El nivel de ingresos incide sobre el uso de internet en actividades académicas.
- El nivel de ingresos incide sobre el uso de internet en actividades de entretenimiento.

b. ¿Incide el uso del internet en el éxito académico que alcanzan los estudiantes de nivel superior?

### Hipótesis

- El uso del internet en actividades académicas incide en el éxito académico de los estudiantes de nivel superior.
- El uso de internet en actividades de entretenimiento inciden en el éxito académico de los estudiantes de nivel superior.

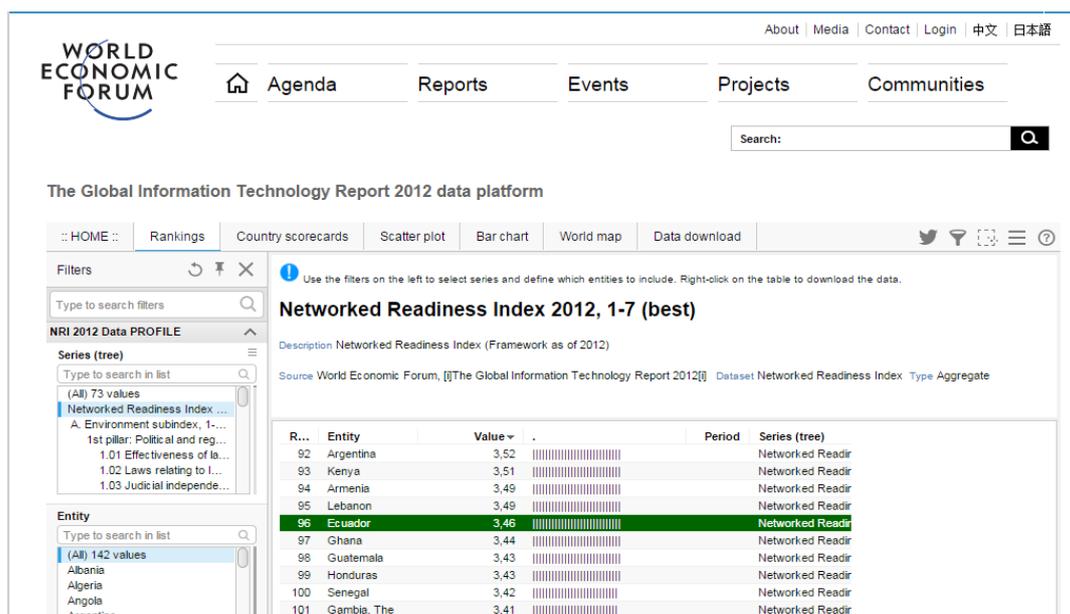
## **CAPÍTULO II - ESTADO DEL ARTE**

## 2. Estado del Arte

### 2.1. Realidad Tecnológica actual en el Ecuador

Los accesos a la tecnología en el Ecuador cada vez son más visibles y tangibles para la mayoría de la población, sin embargo aún existe una brecha digital respecto al nivel alcanzado por otros países.

En el año 2012 en el Foro Económico Mundial (Index & Nri, 2012) llevado a cabo en Nueva York, se presenta a Ecuador en el puesto 96 de 142 países estudiados, esto respecto a componentes de “The Network Readiness Index” que está relacionado a la disposición para aprovechar las TIC’s. En este estudio emprendido por FEM se ve aun la existencia de la brecha digital entre los países desarrollados y los países en desarrollo, siendo Ecuador parte del segundo grupo mencionado [Ver Figura 1].



**Figura 1. The NRI Ranking FEM 2012 (142 países evaluados).**

Fuente: (Index & Nri, 2012)

Este mismo ejercicio ha sido realizado desde el 2008, donde Ecuador ocupaba la posición número 116 de 134 países estudiados (Dutta, 2009), hasta el último año que terminamos que es el 2014 en donde Ecuador ya alcanza el puesto 82 de 148 países estudiados (WEF, 2014). Lo que indica claramente que en los últimos seis años Ecuador ha ganado positivamente 34 posiciones de entre los países estudiados,

logrando con ello superar a países como Argentina, Perú, Venezuela, Guatemala, y a otros países de diferentes continentes [Ver Tabla 1].

Realizando una proyección al 2016 de acuerdo a lo logrado en los últimos seis años (para la proyección de posición se tomó el cálculo del promedio de mejoramiento de posición durante los últimos seis años cuyo resultado es 6, y para la proyección de la puntuación se tomó el cálculo del promedio de mejoramiento de puntuación durante los últimos seis años cuyo resultado es 0.14), Ecuador estaría ubicándose en la posición 70 de los 148 países involucrados en este ejercicio, consiguiendo acercarnos a países como Brasil y Colombia. Ningún país de América Latina ha logrado hasta el momento ubicarse dentro de las primeras 30 posiciones del estudio emprendido por FEM; esto se debe principalmente a la falta de infraestructura tecnológica y desconocimiento de uso de las TIC's por una buena parte de la población [Ver Tabla 1].

**Tabla 1.** The NRI Ranking 2008 - 2014 y Proyección al 2016

AÑO	POSICIÓN	PUNTUACIÓN	MEJOR.POSICIÓN	MEJOR.PUNT	TIPO EJERCICIO
2008	116	3.03	N/A	N/A	EJECUCIÓN
2009	114	3.04	2	0.01	EJECUCIÓN
2010	108	3.26	6	0.22	EJECUCIÓN
2011	102	3.26	6	0	EJECUCIÓN
2012	96	3.46	6	0.2	EJECUCIÓN
2013	91	3.58	5	0.12	EJECUCIÓN
2014	82	3.85	9	0.27	EJECUCIÓN
2015	76	3.99	6	0.14	PROYECCIÓN
2016	70	4.13	6	0.14	PROYECCIÓN

Fuente: Global Information Technology Report (GITR) - Elaboración propia

La Secretaría de Educación Superior, Ciencia, Tecnología e Innovación (SENESCYT) en conjunto con el Instituto de Nacional de Estadística y Censos (INEC) emprendieron un estudio especializado respecto a actividades de ciencia, tecnología e innovación (ACTI) correspondientes al periodo 2009 – 2011 (Senescyt & Inec, 2014), tomando para el mismo una muestra de 161 Instituciones conformadas por: 56 universidades públicas, 13 institutos públicos de investigación, 30 entidades de gobierno, 40 hospitales públicos de docencia y 22 ONG's; de donde se obtuvieron importantes resultados. Se pudo constatar que en el Ecuador en el año 2011 se ha invertido US

\$112'860,000.00 en Investigación y desarrollo, del cual el 74.9% corresponde a investigación aplicada, el 16.4% a investigación básica y el 8.7% a desarrollo experimental, adicionalmente se evidencia que el número de personas dedicadas a la ciencia y tecnología para ese mismo año fue de 3,743 investigadores, de los cuales tan solo el 20.1% estuvo dedicado a la disciplina científica de ingeniería y tecnología, los demás se centraron en otras disciplinas como ciencias sociales, ciencia naturales y ciencias exactas. (Senescyt & Inec, 2014)

Ahora bien, puesto que en el estudio emprendido por la Secretaria de Educación Superior, Ciencia, Tecnología e Innovación, solo ha tomado muestra a entidades públicas o de gobierno, los indicadores expuestos en el párrafo anterior podrían tener un margen de error significativo con tendencia al aumento, pues es evidente que en el Ecuador el sector privado tiene una participación importante de inversión en ciencia y tecnología y esto sin duda aumentaría los indicadores citados; lamentablemente no se ha realizado hasta el momento un estudio similar para instituciones educativas particulares, lo cual imposibilita la comparación de participación que cada uno tiene respecto a la Investigación, Ciencia y Tecnología.

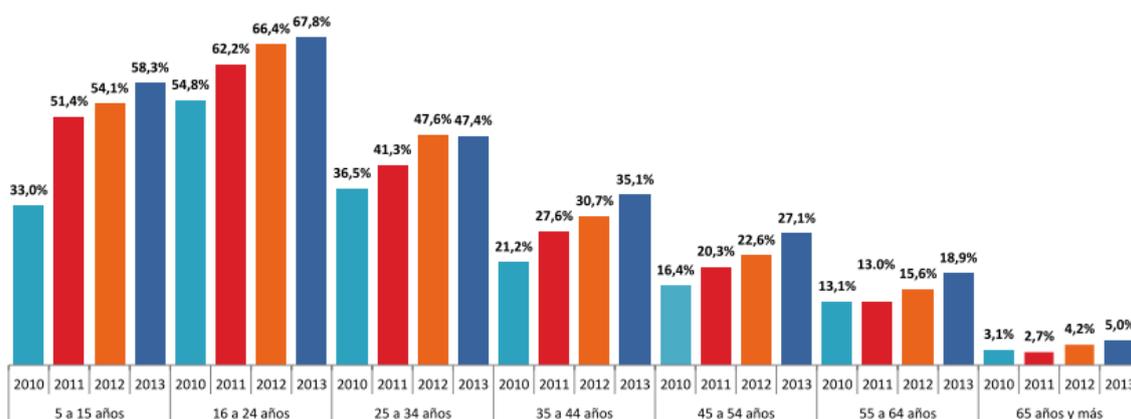
En el Ecuador, tanto el gobierno como el sector privado realizan esfuerzos para dar acceso a la población a las tecnologías actuales existentes a nivel mundial y disminuir la tan nombrada brecha digital; el desconocimiento de las TIC`s agotan esta posibilidad, pues este desconocimiento es causa de que la mayoría de la población carezca de interés de servirse de las mismas (Instituto Nacional de Estadísticas y Censos, 2012).

Otro factor importante dentro de la realidad tecnológica actual del Ecuador, es el alto costo al acceso tecnológico, y de acuerdo a una entrevista realizada por la revista Líderes respecto al tema "convergencia digital" al Consejero del Organismo Regulador de las Telecomunicaciones en España, indica que Ecuador está viviendo de la experiencia que a escala global experimentamos. La suerte de encontrarse en una fase menos desarrollada que otros países, le permite subirse en un tren en marcha y, por lo tanto, no experimentar fases previas de errores cometidos. El proceso convergente es inevitable a escala mundial. Ecuador no puede dar servicios que no estén dentro de esta perspectiva. El despliegue de infraestructura para prestar estos servicios en Ecuador debe ser pensado en la convergencia incluso por una cuestión de política industrial, los equipos para estas redes ya son pensados para la

convergencia. Aquí, lo deseable es que estas inversiones sean eficientes y no redundantes, que se busque la eficiencia y junto a operadores públicos y privados ofrecer servicios a los ciudadanos, para que accedan a los mejores servicios y al más bajo precio posible (Lideres, 2012).

Si bien esta entrevista fue realizada en el 2012, teniendo a la actualidad un poco más de tres años desde su publicación, Ecuador aun no logra servirse plenamente de los beneficios que ofrece la convergencia digital, esto debido a los costos que implica para la población tanto en adquisición de dispositivos que prestan estos servicios, como también en el pago recurrente del servicio como tal.

Uno de los dispositivos tecnológicos más utilizados en Ecuador es la computadora. En el grafico siguiente [Ver Figura 2], se puede evidenciar que de la población total en Ecuador, el grupo con mayor número de personas que utilizan computadora es el que está entre 16 a 24 años con el 67.8%, seguidos del de 5 a 15 años con 58.3%. Con esta información se deduce que el grupo de los jóvenes es el sector que más utiliza la computadora u ordenador, por lo que en este grupo se tendría mayor posibilidad de uso de las TIC's y con ello mayor contribución y participación en la sociedad de la información.

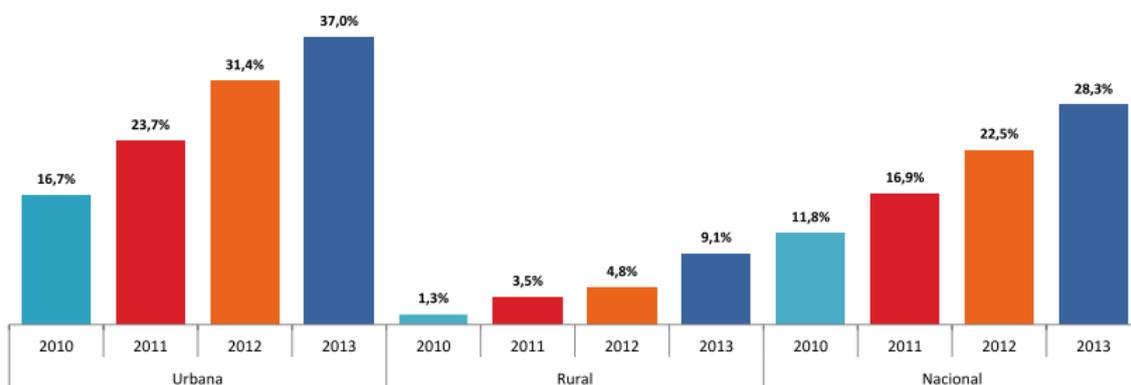


**Figura 2.** Personas que utilizan computadora por grupos de edad en Ecuador.

Fuente: (INEC, 2013)

También respecto al uso de las tecnologías en el Ecuador, (INEC, 2013) nos permite evidenciar que ha ido aumentando la población con acceso a internet y con ello disminuyendo la brecha digital. De acuerdo al informe del INEC, tanto en el área

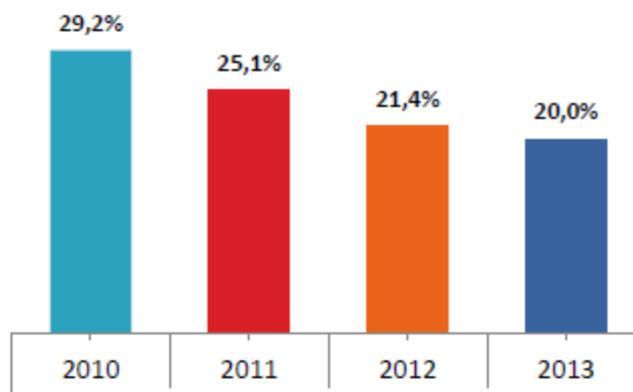
urbana como en la rural ha habido un aumento de los cibernautas. En la [Figura 3], se puede observar que en el 2013 de la población total en Ecuador, tan solo el 28.3% acceden a Internet, de la población que conforma en área urbana el 37.0% acceden a internet, y el 9.1% de la población rural. Con esta información vemos que en Ecuador si bien cada año existe un incremento considerable de acceso a internet, aun existe una desventaja o brecha digital bastante considerable según el estudio de (INEC, 2013).



**Figura 3.** Acceso a Internet según el Área en Ecuador.

Fuente: (INEC, 2013)

Por otro lado, un tema que hace que la brecha digital aun tenga mayor presencia en Ecuador respecto a los demás países a nivel mundial es el analfabetismo digital, sin embargo mediante análisis realizado en (INEC, 2013), en relación al analfabetismo digital latente en Ecuador es necesario comentar que desde el 2010 al 2013 bajo de 29.2% al 20.0% lo que nos indica que si bien en Ecuador tenemos un 20.0% de analfabetismo digital en la población, este porcentaje está decreciendo y con esto disminuyendo la brecha digital; para calificar a una persona dentro del nicho de analfabetismo digital se considera tres características variables: que no tenga celular, que no haya utilizado una computadora en los últimos 12 meses y que no haya utilizado el internet en los últimos 12 meses. Para sustentar lo antes mencionado a continuación se detalla el indicador en la [Figura 4], en donde también se puede evidenciar el decremento de analfabetismo digital en Ecuador desde el 2010 hasta el 2013. Lo cual es positivo respecto a la disminución de la brecha digital



**Figura 4.** Porcentaje de personas Analfabetas Digitales en Ecuador.

Fuente: (INEC, 2013)

## 2.2. Posibilidades de acceso a la educación superior.

En la historia del Ecuador, el acceso a la educación superior muchas de las veces ha estado condicionado por el factor socio-económico de la población. En donde es evidente que la población con recursos económicos limitados está en desventaja respecto a la población de recursos económicos confortables (Morquecho, Erika Priscila Machasilla Piñol & Zeas, 2009). Sin embargo el Gobierno ecuatoriano se ha propuesto diseñar una sociedad del conocimiento, con la característica fundamental de que sea dirigida a todos los estratos de la población, según lo manifiesta (Long, Guillaume, 2014). Informa además que este proyecto busca alcanzar la sociedad del conocimiento, como una vía para erradicar la pobreza, cambiar la matriz productiva, profundizar la democracia y consolidar un sistema de educación progresista y de vanguardia. (Long, Guillaume, 2014). Para lograr lo antes expuesto el Ecuador realiza una inversión para la educación superior es del 2% del producto interno bruto (PIB). También esto ha permitido una mayor inversión en investigación alineados al formar una sociedad del conocimiento.(Diario el Ciudadano, 2014) y (Presidencia-Ecuador, 2013).

El Ecuador está en una etapa de constante evolución educativa, en donde su ambicioso objetivo es lograr que la educación superior esté al alcance de toda la población, para ello existen algunos programas y leyes en las que se establece el derecho a acceder una educación superior de calidad y pertinente, que permita formarse académicamente en igualdad de oportunidades (Presidencia de la Republica,

2010); con ello se pretende romper la barrera socio-económica que genera una desventaja a quienes tienen menor posibilidad económica.

### **2.3. Evolución universitaria desde una perspectiva Tecnológica.**

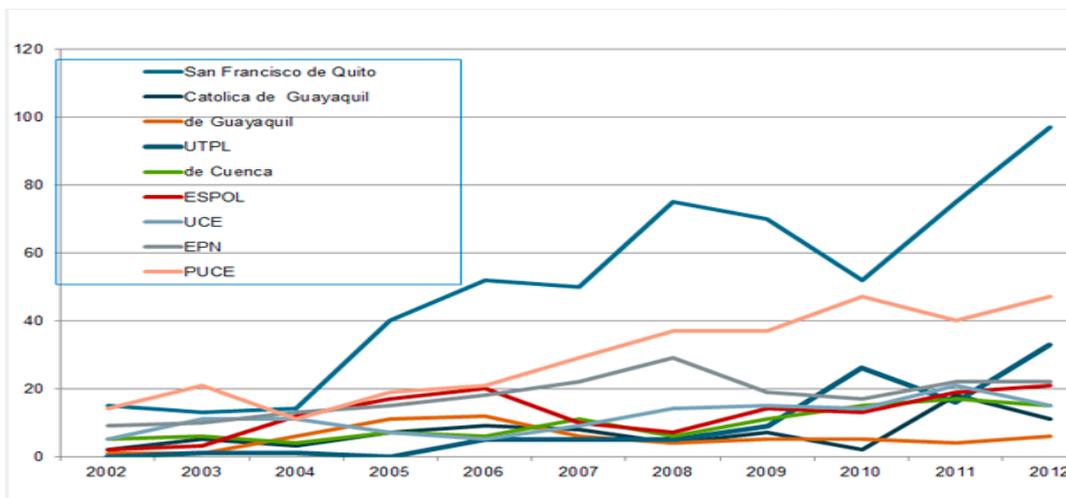
La Universidad desde sus inicios por su naturaleza está orientada a la innovación. La importante evolución que han tenido las TIC's y dentro de ellas el internet, ha abierto la posibilidad para dar un giro a la educación superior (Coll & Martí, 2007). Innovaciones como la web 2.0 han logrado implementar en la educación superior un modelo de colaboración, en donde los colaboradores pueden ser docentes o estudiantes, a través de diferentes herramientas como foros, wikis, redes sociales, blogs, aulas virtuales, etc (Francisco Amat, 2011). Otra innovación importante es la convergencia digital, que permite acceder y compartir información desde diferentes dispositivos también innovadores como: Computadoras, tabletas, smartphones, televisores, y otras herramientas smart (Soto Arango, Mesa Jiménez, & Orlando Caro, 2012).

El internet, constituye el engranaje maestro en donde todas las tecnologías innovadoras se apoyan para poder funcionar y brindar su capacidad de servicio. Así, en la educación superior tanto ésta, como otras tecnologías inciden de manera significativa dentro y fuera del salón de clases. Consolidando las razones o necesidades que nos llevan a usar las TIC's en la educación tenemos: la alfabetización de los alumnos, la Productividad que genera tanto para docentes como para estudiantes, y la innovación en las practicas docentes (Marqués Graells, 2012).

Las Universidades con la ayuda de las TIC's han podido innovar en su forma de enseñar, permitiendo que sea posible el estudio a distancia, que amplía las posibilidades de acceso a la educación superior, y es utilizado principalmente por la población que ya tiene un empleo. Este tipo de educación es de tipo virtual, por lo que tanto profesores como alumnos deben dar un giro a la forma de enseñanza tradicional y abrirse hacia metodologías de aprendizaje con más posibilidades y mayor flexibilidad (Hinojo & Fernández, 2012).

Adicionalmente las tecnologías han permitido a las Universidades ampliar sus líneas de investigación y alcance a nivel internacional, y un indicador claro es el crecimiento en la indexación de publicaciones científicas. La mayoría en la base de datos de

Scopus que es hoy en día la mejor herramienta para estudios bibliométricos y evaluaciones de producción científica. De acuerdo a gráficos publicados en la sección de investigación del portal Web de la UTPL la publicación de artículos científicos en las universidades del Ecuador ha ido creciendo considerablemente en estos últimos años [Ver Figura 5].



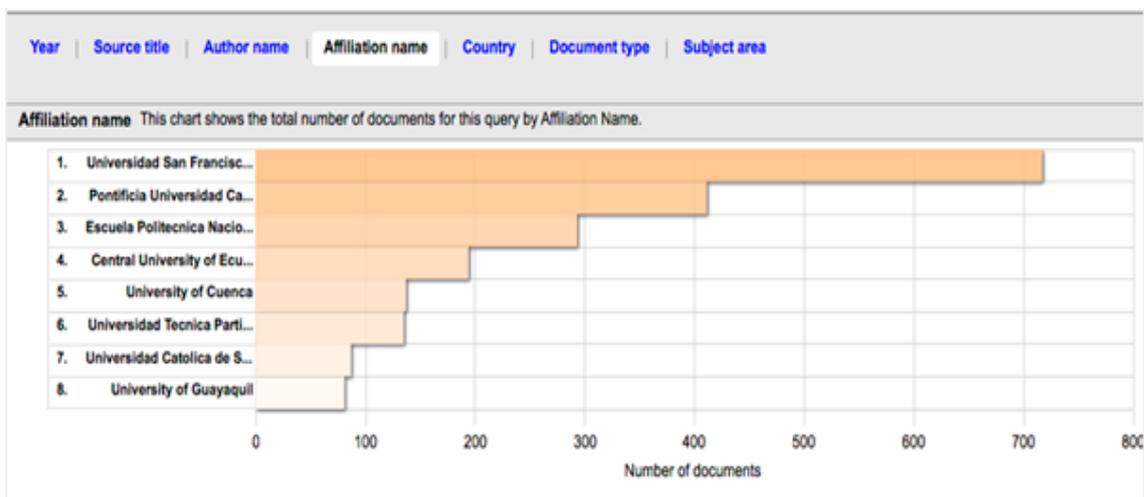
**Figura 5.** Número de publicaciones en las Universidades de Ecuador hasta el 2012.  
Fuente: (UTPL, 2013)

De acuerdo al número de artículos publicados acumulados en el tiempo en base de datos Scopus con corte al 2012, las ocho universidades del Ecuador que más publicaciones han tenido realizado son (UTPL, 2013).

- Universidad San Francisco de Quito
- Pontificia Universidad Católica del Ecuador
- Escuela Politécnica Nacional.
- Universidad Central del Ecuador.
- Universidad de Cuenca.
- Universidad Técnica Particular de Loja
- Universidad Católica de Guayaquil.
- Universidad de Guayaquil.

Lo descrito en líneas anteriores puede también visualizarse en la [Figura 6], en donde se puede constatar que la Universidad San Francisco de Quito es de las Universidades del Ecuador la que más ha aportado con publicaciones científicas. En los primeros 10 meses del 2013 las universidades del Ecuador alcanzaron 500

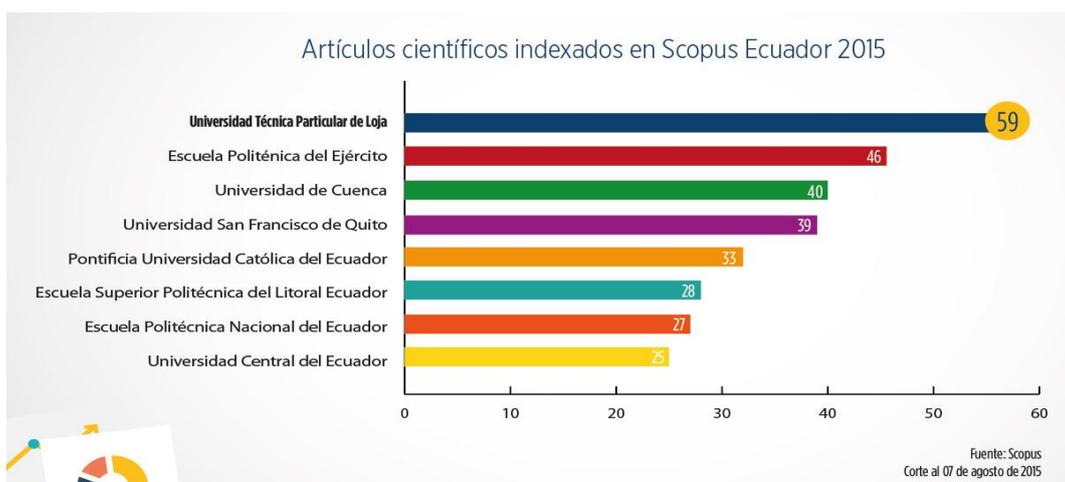
publicaciones indexadas en Scopus (Bruque, 2013) cifra que no se había logrado antes.



**Figura 6.** Número de publicaciones en SCOPUS Universidades de Ecuador.

Fuente: (UTPL, 2013)

Sin embargo un nuevo reporte de cifras generado en el 2015 muestra un nuevo número de artículos indexados en la base de datos Scopus en el año 2015, con corte al 07 de Agosto, en el cual se puede evidenciar que si bien algunas Universidades siguen siendo las mismas que despuntan en producción científica, existen otras que ya no aparecen entre las primeras ocho. También algo que hay que indicar es que el lugar en que se ubicaban al año 2013 dichas instituciones ha variado al 2015 [Ver en la Figura 7].



**Figura 7.** Número de publicaciones en SCOPUS Universidades de Ecuador en el 2015.

Fuente: (UTPL, 2015)

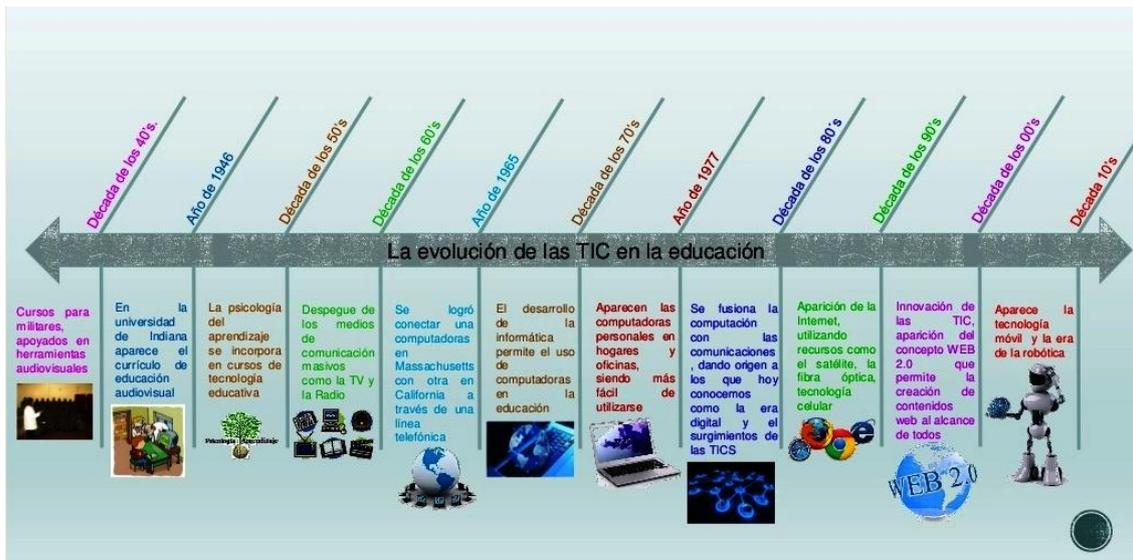
Entonces con este nuevo reporte que se presenta en orden de acuerdo al número de artículos publicados en base de datos Scopus en el transcurso del 2015 hasta el 7 de Agosto, las ocho universidades del Ecuador que más publicaciones han tenido realizado son: (UTPL, 2015).

- **Universidad Técnica Particular de Loja**
- Escuela Politécnica del Ejército.
- Universidad de Cuenca.
- Universidad San Francisco de Quito
- Pontificia Universidad Católica del Ecuador,
- Escuela Superior Politécnica del Litoral Ecuador
- Escuela Politécnica Nacional del Ecuador.
- Universidad Central del Ecuador.

En este nuevo reporte la Universidad que ocupa el primer lugar en producción científica en el 2015 es la Universidad Técnica Particular de Loja con 59 artículos indexados en SCOPUS, logrando con ello subir del sexto lugar al primer lugar en estos últimos dos años.

#### **2.4. La Web y el impacto que ha ocasionado en la Educación Superior.**

La Web ha dado pasos agigantados en su evolución y desarrollo, hace algunos años el número de sitios web no era representativo y los mismos estaban compuestos solamente de texto; si bien ya desde entonces la web inicia su relación con el ámbito educativo superior, el uso de la misma era limitada tanto por su difícil acceso, como por su baja funcionalidad. Pero la evolución de la tecnología también ha permitido evolucionar la forma de programar y administrar los sitios web, pasando de páginas estáticas y sin interacción directa con el usuario a páginas dinámicas que administran mucha información de los usuarios, produciendo con ello el surgimiento de la Web 2.0 que dio paso a la formación de la comunidad virtual, esto puede apreciarse en la **[Figura 8]**. (Lemus, 2014).



**Figura 8.** La evolución de las TIC en la educación.

Fuente: (Lemus, 2014)

(Rittberger & Blees, 2009) explican la relación de la web 2.0 y el aprendizaje, y plantea algunas interrogantes que deberíamos tenerlas claras; Web 2.0 es un salto cualitativo en la tecnología web que han hecho que internet sea más creativo, participativo y socializante, pero ¿Acaso este desarrollo ha supuesto una revolución en el aprendizaje? ¿Acaso requiere la educación y el aprendizaje un replanteamiento con vistas a los cambios continuos de las tecnologías de información y comunicación? ¿Necesitamos nuevos conceptos diseñados para los entornos de trabajo y aprendizaje respectivamente?. Los instrumentos de web 2.0 se convierten en relevantes rápidamente como causa de los intercambios, más allá del conocimiento y del desarrollo de las competencias en las redes y más allá de la red de una forma óptima.

Con la aparición de la Web 2.0 el internet ha pasado a ser una plataforma de creación y administración de contenido, permitiendo reutilizar y compartir la información en tiempo real, y cambiando el papel del usuario de ser un lector a ser un escritor-administrador de páginas web. Podemos dividir en cuatro formas de aprendizaje que la web 2.0 permite a los usuarios (Gil-juarez, Vitores, Feliu, & Vall-llovera, 2011):

- Aprender haciendo [learning-by- doing]: Para este tipo de aprendizaje resultan de especial utilidad aquellas herramientas que permiten al estudiante y/o docente la lectura y la escritura en la Web, bajo el principio de “ensayo-error”. Por ejemplo, los estudiantes pueden aprender sobre ecología generando presentaciones en línea (de texto, audio o vídeo) sobre cómo se trata este

tema en diferentes naciones del mundo. Luego el profesor lo revisa y corrige aquellos aspectos mejorables. Este proceso de creación individual y colectivo a la vez, promueve un proceso de aprendizaje constructivista.

- Aprender interactuando [learning-by-interacting]: Una de las principales cualidades de las plataformas de gestión de contenidos es que además de estar escritas con hipervínculos, ofrecen la posibilidad de intercambiar ideas con el resto de los usuarios de Internet. Bajo este enfoque, el énfasis del aprender interactuando está puesto en la instancia comunicacional entre pares. Algunos ejemplos de interacción son: agregar un post en un blog o wiki, enviar un voice mail y actividades tan coloquiales como usar el chat, el correo electrónico o las redes sociales.
- Aprender buscando [learning-by-searching]: Uno de los ejercicios previos a la escritura de un paper, trabajo, ensayo o ejercicio, es la búsqueda de fuentes que ofrezcan información sobre el tema que se abordará. Ese proceso de investigación, selección y adaptación termina ampliando y enriqueciendo el conocimiento de quien lo realiza. En un entorno de gran cantidad de información disponible, resulta fundamental aprender cómo y dónde buscar contenidos educativos.
- Aprender compartiendo [learning-by-sharing]: El proceso de intercambio de conocimientos y experiencias permite a los educandos participar activamente de un aprendizaje colaborativo. Tener acceso a la información, no significa aprender: por esto, la creación de instancias que promuevan compartir objetos de aprendizaje contribuyen a enriquecer significativamente el proceso educativo. Internet cuenta con una gran cantidad de recursos para que los estudiantes puedan compartir los contenidos que han producido. Por ejemplo: plataformas para intercambio de diapositivas en línea, podcasts o videos educativos.

La Web 2.0 ha permitido tomar nuevos horizontes dentro de la educación superior, siendo principal responsable de orientar a la sociedad a ser una sociedad de la información. Para vivir, aprender y trabajar con éxito en una sociedad cada vez más compleja, rica en información y basada en el conocimiento los estudiantes y docentes deben utilizar la tecnología digital con eficacia. En un contexto educativo sólido, las Tecnologías de la Información y la Comunicación (TIC) pueden ayudar a los

estudiantes a adquirir las capacidades necesarias para llegar a ser: competentes para utilizar tecnologías de la información; buscadores, analizadores y evaluadores de información; solucionadores de problemas y tomadores de decisiones; usuarios creativos y eficaces de herramientas de productividad; comunicadores, colaboradores, publicadores y productores; y ciudadanos informados, responsables y capaces de contribuir a la sociedad (Francisco Amat, 2011).

Uno de los actores principales de la educación superior es el docente o educador, y la evolución de la web también ha hecho que dichos actores cambien el rol pedagógico que cumplen en la universidad, pasando de ser un docente tradicional a ser prosumidor (productor - consumidor) de los estudiantes o educandos y pares académicos de los mismos (K. Guerrero & Caballero, 2013), sin dejar de lado una evaluación consciente y autocrítica para mejorar los ambientes de aprendizaje.

Un motor importante dentro de la sociedad de la información y de la educación superior como tal es la “convergencia digital”, misma que ha sido posible por la evolución de la web y su espacio digital. La convergencia digital permite trabajar simultáneamente con información desde diferentes dispositivos tecnológicos que cumplen la función de habilitantes. Lo que facilita a los docentes y a los estudiantes compartir y acceder a información para su día a día en los salones de clase y en los espacios dedicados a la investigación como parte de su proceso de aprendizaje (Soto Arango et al., 2012). La fórmula: *convergencia de medios + convergencia expresiva = (Cooperación Internacional + responsabilidad social + solidaridad) / identidad = Sociedad del Conocimiento – Brechas Digitales y Sociales = Sociedad del desarrollo mundial*, es una fórmula que nos puede hacer conquistar el futuro (García García, 2012)

## **2.5. La brecha digital en la sociedad de la información.**

En el mundo, la información se ha convertido en la corona principal del engranaje que mueve todos los campos públicos y privados, tales como: la economía, la educación, la ciencia, la salud, la comunicación, y la más importante que a su vez engloba todas las anteriores que es la investigación. Por otro lado al ser un mundo inundado de datos, es responsabilidad de las instituciones educativas el enseñar a interpretarla, organizarla y entenderla con una visión de 360 grados, con la finalidad de lograr ciudadanos activos y críticos ante la sociedad del siglo XXI, y esto recae directamente sobre el rol del docente (Francisco Amat, 2011).

Ahora, si bien nos encontramos viviendo en una sociedad de la información misma que cobra sentido cuando sirve de medio para fines más elevados y deseables que concluyen en la formación de la sociedad del conocimiento, el acceso a la información ha originado una desigualdad o brecha digital, misma que pone en desventaja la forma de aprendizaje actual de los estudiantes, y esto repercute directamente en el rendimiento académico de los mismos. La desigualdad o brecha digital se puede medir desde dos temas macro:

- a. Acceso a la información.\_ Comprende la posibilidad o limitación que tienen los usuarios para acceder a la información por medio de las herramientas actuales y de vanguardia. Si bien el número de usuarios que tienen acceso a internet es bastante, aún existe un porcentaje significativo de la población que no tiene acceso a internet, esto por su ubicación, desconocimiento o por su nivel socio-económico actual, esto hace que la brecha digital disminuya o aumente de acuerdo a la incidencia del número de usuarios que no tienen acceso a internet. De acuerdo al estudio emprendido por el instituto de nacional de estadística y censo del ecuador (INEC, 2013) tan solo el 40.4% de la población tiene acceso a internet, de este porcentaje de la población el 64% usa el internet con una frecuencia diaria, 32.7% con una frecuencia semanal y el 3.3% con una frecuencia mensual o anual. Por otro lado a nivel de América Latina y el Caribe el número de habitantes que tienen acceso a Internet es del 46.72% (A4AI, 2013) (Rojas, Poveda, & CEPAL, 2015); y a nivel mundial el número de habitantes que tienen acceso a internet es el 43.4% (De & Telecomunicaciones, 2015). Tomando estos datos de acceso a internet, el Ecuador tiene una brecha digital respecto al acceso a internet del 6.32% respecto a la población de América Latina y tan solo al 3% respecto a la población a nivel mundial.
- b. Uso de información.\_ Trata sobre los diferentes usos que se le puede dar a la información por parte de los usuarios. A pesar de muchos usuarios tener acceso a la información, los usos que la dan a la misma no siempre son los adecuados o no van en línea a la disminución de la brecha digital. De acuerdo a encuestas emprendidas por (INEC, 2013) a la población ecuatoriana en el 2013, del universo de usuarios que tienen acceso a internet el 32% de usuarios utilizan el internet como fuente de información, el 31.7% utilizan como educación y aprendizaje y el otro 36.3% lo utilizan para otros temas varios

como comunicación en general que no van enfocados directamente a la autoeducación de los usuarios. No se ha logrado obtener información a nivel mundial, lo cual impide el análisis de porcentajes de usos de internet a nivel mundial.

Lo descrito en los literales a y b de esta sección nos permiten visualizar fácilmente la existencia de la brecha digital en la sociedad de la información. La disminución de dichas brechas depende mucho de las instituciones educativas, pues como bien lo mencionan (Gil-juarez et al., 2011) en su investigación “Teoría de La Educación Y Cultura En La Sociedad de La Información” si bien se ha logrado una mayor frecuencia de acceso a las TIC’s por parte de los estudiantes, los mismos muestran una comprensión limitada de sus necesidades de información y presentan dificultades en búsquedas efectivas, lo que provoca una falta de habilidades críticas y analíticas para juzgar la relevancia y la confiabilidad de lo que encuentran en el internet.

Con el apoyo de un sistema de educación robusto orientado a las tecnologías de la información y la comunicación (TIC) se puede lograr que los estudiantes adquieran capacidades importantes que los lleve a convertirse en competentes dentro de una sociedad de la información y del conocimiento, analizadores y evaluadores de la información, solucionadores de problemas y emprendedores de innovaciones, responsable y capaces de contribuir a la sociedad actual.

## **2.6. El rendimiento o éxito académico frente al uso de internet.**

Existen tres niveles o definiciones que ocasionan la “brecha digital” según (Norris, 2001); la de *nivel global* que se refiere a la desigualdad entre los países desarrollados y los que se encuentran en proceso de desarrollo, la de *nivel social* que hace referencia a la desigualdad en una sociedad o nación respecto a quienes tiene acceso a la información y quienes carecen de él, y la de *nivel democrático o individual* que es la brecha que existe entre las personas que cuentan con acceso a las TIC, pero el mismo es aprovechado de forma distinta según la habilidad, experiencia, educación y soporte social con la que cuenta cada usuario o persona (A. R. Gallardo, 2006); este último nivel de brecha también es conocida como “inequidad digital” (DiMaggio & Hargittai, 2001).

En línea a lo expuesto concierne analizar la relación entre los diferentes usuarios de internet y el desempeño académico de los mismos, como una forma clara de ver la interacción en el desarrollo del usuario respecto a estas nuevas oportunidades informáticas; para lo cual dentro del universo de los usuarios los mas optados para un resultado efectivo de este trabajo de investigación son los estudiantes universitarios, por la diversidad de personas que encontramos en este grupo.

Por rendimiento o éxito académico se entiende al resultado obtenido por cada estudiante en el año de estudios, ya sea este resultado positivo o negativo. Pues por otro lado el uso del internet puede ser muy diverso entre los usuarios, pero para este trabajo de investigación se lo procederá a segmentar en dos grupos por su forma de uso: La primera sería la forma positiva de sacarle provecho a las bondades que nos brindan las nuevas tecnologías aplicándola en el aprendizaje que en la actualidad es de vital importancia para la educación superior. Respecto a este tema se han realizado algunos trabajos de investigación, encontrando resultados positivos y negativos al respecto, entre algunos de ellos tenemos el realizado por (Torres, 2012) que sugiere el 3% de incidencia positiva, (Castaño Muñoz, 2011) sugiere el 3.12% de incidencia positiva respecto a Universidades presenciales y 4.97% de incidencia positiva en Universidades en Modalidad a Distancia o Virtual. Y la segunda sería la forma de usar las nuevas tecnologías como pleno entretenimiento y pasatiempo, que al final se convierte en un distractivo para las actividades académicas, que de acuerdo a algunos trabajo de investigación apuntan a resultados positivos y negativos respecto al éxito académico; entre algunos de ellos tenemos el realizado por (Torres, 2012) que sugiere el 1.9% de incidencia positiva. En (Frangos, Frangos, & Kiohos, 2010) se llegó a la conclusión que la utilización del internet para entretenimiento tiene un efecto negativo respecto a éxito académico del estudiante, toda vez que este uso de internet provoca la distracción del estudiante en sus estudios.

La eficiencia de la utilización del internet para el aprendizaje aumenta el éxito académico de acuerdo a la intensidad con la que se la utilice, y este aumento tiene un límite a partir del cual no se puede obtener más beneficios (Castaño Muñoz, 2011).

La mayoría de países en América latina han apostado proyectos interesantes respecto a la universalidad de acceso a internet hacia los centros educativos con la finalidad de reducir la brecha digital en los estudiantes, y en estudios emprendidos por estos mismos países posteriormente han arrojado resultados que apuntan a una alta

penetración de las TIC's en los centros educativos, pero así mismo dejan ver en muchos alumnos su inadecuado aprovechamiento.(López, Gonzalez-barrera, & Patten, 2013).

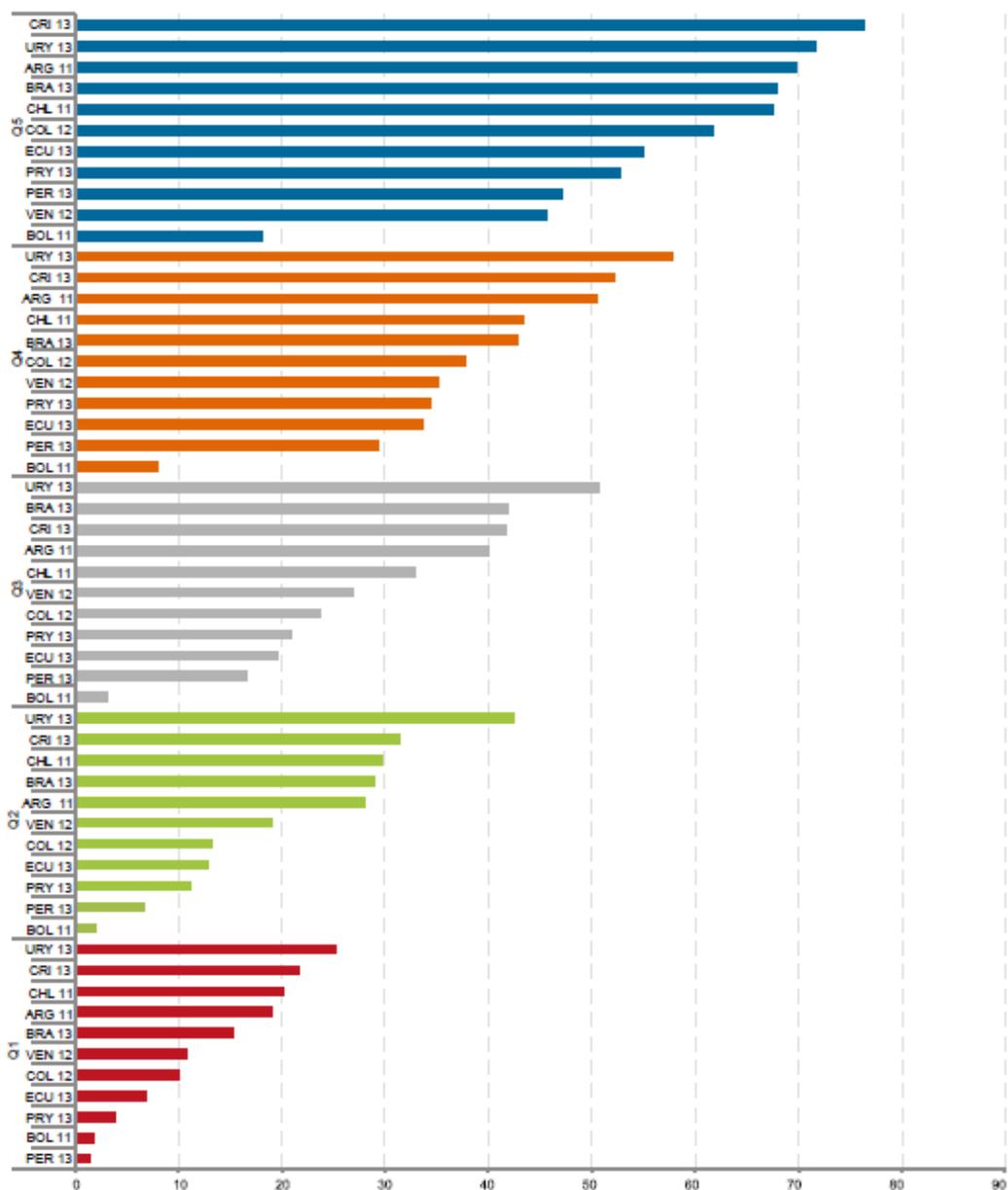
## **2.7. El nivel socioeconómico frente al uso de internet.**

Uno de los factores que influye ampliamente a la formación de la brecha digital en los países en proceso de desarrollo es el nivel socioeconómico de la población, pues en estos países el costo de acceso a internet aun tiende a ser elevado, y esto influye directamente a los jóvenes estudiantes que se están preparando en las aulas universitarias.

El nivel socioeconómico marca una desigualdad de uso de internet en aquellos estudiantes que tienen menor estatus socioeconómico, pues ellos suelen incidir en emociones negativas al momento de interactuar con el ordenador, lo cual ocasiona que utilicen menos el internet (Castaño-Muñoz, 2010).

El nivel socioeconómico está ligado directamente al analfabetismo digital, pues muchas personas por sus escasos ingresos no tienen la posibilidad de tener un computador, o uno de los dispositivos tecnológicos que les permita tener acceso a internet.(Cusí, 2003)

En los estudiantes universitarios es común ver que muchos de ellos prefieran acceder al internet desde el campus universitario que desde su hogar, esto puede darse por múltiples razones, pero una de ellas es debido a que muchos no tienen acceso a internet en sus hogares porque su economía no se los permite, lo cual les genera una desventaja en relación a estudiantes con mejores posibilidades económicas. La Comisión Económica para América Latina y el Caribe de acuerdo a estudios emprendidos en el 2015 nos muestra mediante grafica en quintiles [Ver Figura 8], la desventaja que tienen los hogares más pobres (Q1) frente a los hogares más ricos (Q5) respecto al acceso de internet. Por ejemplo en Ecuador se puede evidenciar que en el Quintil Uno que corresponde a los hogares más pobres alcanza alrededor del 7% de acceso a internet, mientras que en el Quintil Cinco que corresponde a los hogares más ricos alcanzan alrededor de 55% de acceso a internet, por lo que la brecha digital entre estos dos Quintiles sería del 48% que corresponde a la brecha digital en Ecuador por el factor económico en las familias.



**Figura 9.** Hogares con acceso a Internet según quintil de ingresos.

Fuente: (Rojas et al., 2015)

Adicional a la relación que tiene el nivel socioeconómico y el acceso a internet, lo cual ya pudimos observar mediante la [Figura 8], también se han realizado estudios en Ecuador respecto a la relación que tiene el nivel socioeconómico con el uso que cada estudiante puede dar al internet. Para este trabajo de investigación se ha considerado dos formas de uso de internet; la primera que se refiere a la incidencia que tiene el nivel socioeconómico de los estudiantes frente al uso del internet para actividades académicas, para el cual mediante estudio realizado por (Torres, 2012) nos refiere a un 9% de incidencia. Y el segundo que se refiere a la incidencia que tiene el nivel

socioeconómico de los estudiantes frente al uso del internet para actividades de entretenimiento, en donde mediante estudio realizado por (Torres, 2012) nos refiere a un 16.9% de incidencia; es decir los estudiantes que tienen menos nivel de ingresos, tienen a interactuar menos en internet para actividades de entretenimiento.

Como resultado de investigación emprendida por (Torres e Infante, 2011), el nivel de ingresos familiares genera una brecha respecto al uso del internet, pues los estudiantes con mayores ingresos familiares están sobre el perfil alto del uso de internet, mientras que los estudiantes con ingresos económicos bajos se encuentran sobre el perfil bajo de uso del internet.

### **CAPÍTULO III - MARCO TEORICO**

### 3. Marco Teórico

#### 3.1. La brecha digital

En esta última década el Internet se ha convertido en una necesidad, facilitando que de una manera fácil pasemos de ser una sociedad convencional a ser una sociedad digital o de la información, provocando con ello colateralmente el inicio de una desigualdad digital o también conocida como *digital divide* de la población, en el ámbito social, económico, educativo, en fin en todos los ámbitos que globalmente incluye la sociedad actual. Muchos investigadores se han interesado en estudiar las causas y efectos que producen esta desigualdad digital, como (Castaño-Muñoz, 2010) que nos indica que la investigación sobre la *Digital Divide* ha puesto en manifiesto el cómo no solo es importante estudiar las diferencias entre la gente que tiene acceso a internet y la que no. Actualmente existen otras dimensiones que cabe estudiar: habilidades en el uso de Internet, tiempo en la red y especialmente los tipos de usos que la gente hace de Internet. Igualmente, para cada una de estas dimensiones es importante estudiar sus determinantes y sus consecuencias.

(López, 2007) en su investigación sobre la influencia de la Tecnología de la información en la educación, nos comenta que con el inexorable transcurrir del tiempo, uno comienza a entrever, de una forma incipiente, e incluso algunas veces primitiva, la complejidad que subyace en el momento histórico que nos ha tocado vivir, un espacio temporal caracterizado por el descubrimiento de una plétora de medios de comunicación de masas (cine, prensa, radio, televisión digital, videojuegos online, teléfonos móviles con cámara digital y grabador/hologramas, reproductores y grabadores de imágenes digitales y la gran estrella, la red Internet) que conviven en las formas de comunicarnos y en los estilos y modos de consumir nuestro tiempo de ocio.

También (Castaño Muñoz, 2011) resume en su investigación respecto al “Uso de Internet para la interacción en el aprendizaje” explicándonos que sobre la base de una concepción social de la tecnología, podemos afirmar que la tecnología se adopta en diferentes formas y con diversos fines por los usuarios. El Internet en la educación no es una excepción, y los diferentes usos de esta tecnología por los estudiantes pueden llevar a desigualdades en el rendimiento académico, en algunos de los casos, estos pueden deberse a una continuación de "malos usos" que ya existía antes de que el

estudiante comenzara a utilizar el Internet, mientras que en otros casos estos pueden ser debido a obstáculos institucionales para la adaptación de las nuevas formas de enseñanza a las nuevas formas de aprendizaje.

En estudio emprendido por (Pick & Azari, 2008). Respecto a “Global digital divide: Influence of socioeconomic, governmental, and accessibility factors on information technology” concluye que mientras los factores tecnológicos en los países desarrollados están fuertemente asociados a publicaciones científicas, seguida de inversión extranjera directa, porcentaje de las mujeres en la fuerza laboral, y las variables educativas; en los países en desarrollo los factores tecnológicos están fuertemente asociados con la inversión extranjera directa, y el establecimiento de prioridades por parte del Gobierno en Tecnología y las variables educativas.

Con lo expuesto en los párrafos anteriores respecto a los estudiosos que han emprendido investigaciones con relación a la brecha digital “Digital Divide”, se puede concluir que la brecha digital depende de algunos factores relacionados, tanto de la posibilidad de accesos a la TIC's como el uso que los usuarios les dan a las mismas, y en el caso de nuestro estudio, esto incide directamente en desigualdades en el rendimiento académico de los estudiantes universitarios o de nivel superior.

### **3.2. Factores que influyen en la brecha digital.**

Los factores que influyen en la formación y aumento de la brecha digital a nivel mundial no actúan aisladamente, sino más bien se combinan entre sí de acuerdo a la situación que vive cada nación o localidad. Es por ello que la brecha digital se la debe considerar como una situación sensible y de frágil impacto para la sociedad, y que es mucho más propensa a aumentar en los lugares que son más vulnerables a presentar los siguientes factores que influyen en la brecha digital de acuerdo ha informe “Hacia la sociedad del conocimiento” de (UNESCO, 2005):

#### **3.2.1. Los ingresos económicos.**

El costo de acceso a internet y a ordenadores todavía tiende a ser elevado para los estudiantes y usuarios en general de algunos de los países en proceso de desarrollo, y es por ello que estos puntos representan un poderoso factor de desigualdad. Una inversión de infraestructura realizada por el organismo gubernamental y el subsidio de

un porcentaje del costo del servicio de acceso a internet para la población en general lograría disminuir la brecha existente por este factor.

### **3.2.2. La edad.**

Dentro de la población existen miembros de todas las edades, entre ellos están:

- Los niños, que desde su corta edad ya tienden a involucrarse con la tecnología e información digital. Este sector de la población es considerado el futuro de una sociedad de la información.
- Los jóvenes, que inciden en la utilización de las innovaciones tecnológicas y sus aplicaciones, sin embargo estos son vistos como participantes vulnerables a las dificultades económicas y sociales. Este sector de la población es considerado como los educandos de una sociedad de la información.
- Los adultos, que son los protagonistas actuales y responsables de la construcción de una sociedad de la información, dentro de este sector encontramos a los docentes, directivos, profesionales independientes, empresarios, ejecutivos, entre otros. Por otra parte, si bien muchos de los participantes que conforman este sector son facilitadores y promotores dentro de la sociedad de la información, existe un nicho de participantes de cierta edad que no logran ir al ritmo actual de las innovaciones tecnológicas y esto puede ser un obstáculo insuperable dentro de la sociedad de la información que incide directamente en la formación de la brecha digital. Este sector de la población es considerado como los educadores y promotores de una sociedad de la información.

Consolidando los tres sectores según las edades del total de la población, es evidente la apreciación de la importancia y responsabilidad que tienen los mismos para la contribución de la sostenibilidad de una sociedad de la información y del conocimiento. Una formación sistemática de los niños y jóvenes en las tecnologías innovadoras y una mutua solidaridad entre las generaciones de participantes para el beneficio común, permitirían reducir las brechas existentes y contribuirían permanentemente a reforzar los vínculos dentro de una sociedad del conocimiento.

### **3.2.3. El Género.**

Las desigualdades entre hombres y mujeres en el ámbito de las nuevas tecnologías es otra faceta de la brecha digital. En efecto, casi los dos tercios de los analfabetos del mundo son mujeres. En los países en desarrollo, una mujer de cada dos por término medio no sabe leer. Aunque en los países industrializados las mujeres representan una proporción bastante considerable de los usuarios de Internet, en los países en desarrollo existe el riesgo de que acumulen una serie de desventajas que les impidan acceder a las nuevas tecnologías.

Sin embargo es necesario tomar en cuenta que esta amplia brecha digital que se menciona en el informe de UNESCO con relación al género descrita en el párrafo anterior, ya ha disminuido en estos últimos años, sobre todo es más palpable en los países en proceso de desarrollo, pues la mujer poco a poco ha ido cubriendo esa desigualdad social que se venía marcando por la cultura e historia de los pueblos de todo el mundo.

(Hargittai, 2010). En su estudio *Digital Natives? Variation in internet skills and uses among members of the "net Generation."* obtuvo dentro de los resultados de la investigación que las mujeres afirman los niveles más bajos de conocimientos en relación al uso del Internet, pero la diferencia estadísticamente es mínima.

Un motivo adicional por el que aun mantiene la desventaja tecnológica la mujer frente al hombre es por las actividades que realiza cada uno dentro de la sociedad actual, sobre todo en los países en desarrollo. Pues mientras el hombre pasa la mayoría del día en su trabajo utilizando la tecnología, y por las responsabilidades de su trabajo muchas de las veces sigue conectado en casa mediante dispositivo Smartphone; una gran cantidad de mujeres se quedan en sus hogares realizando actividades tradicionales y propias del hogar, cuyo tiempo empleado en estas actividades es la medida de desventaja frente al de su género contrario.

Otros investigadores tales como (Howard & Massanari, 2007), y (Cooper, 2006) coinciden en la afirmación de que la brecha digital respecto al factor género es evidente y se lo ha tenido desde que entro en auge la tecnología en la sociedad. Por otro lado también evidencian que dicha brecha esta disminuyente por los cambios socio-culturales que se han venido dando en el transcurso del tiempo a nivel mundial.

#### **3.2.4. El idioma.**

La diversidad de lenguas que se aplican en los diferentes países del mundo, mismas que por naturaleza también se aplican cuando los participantes de cada país publican sus investigaciones y aportes, tiende a limitar el acceso a información valiosa a las personas de países que ejercen un idioma distinto, apuntando hacia un crecimiento de la brecha digital, actualmente existen dos idiomas vectores de publicaciones en internet que son: inglés y español, también existen muchas lenguas que casi no se utilizan en el internet, lo cual margina a las personas que practican las mismas, frente a quienes practican las lenguas vectores descritas en líneas anteriores. Una posibilidad de disminución de esta desventaja es la implementación de herramientas de traducción simultánea, que actualmente ya existe en muchas de las páginas de Internet, de al menos los idiomas más hablados en el mundo, lo cual ha ayudado a la disminución de la brecha digital provocada por la diversidad de idiomas.

#### **3.2.5. La geografía.**

La asimetría entre ciudades y los múltiples poblados rurales crean sin lugar a duda la desigualdad tecnológica por algunas causas, entre ellas la limitación de acceso a internet por su situación geográfica. El movimiento económico en las ciudades hace que las empresas de telecomunicaciones inviertan mucho más capital en los servicios que ofrecen en el sector urbano, dejando en desventaja tecnológica al sector rural. Sin embargo en estos últimos años la competencia entre las empresas de telecomunicaciones ha hecho que amplíen sus servicios hacia el sector rural, lo que ha apoyado evidentemente a la disminución de la desigualdad o brecha digital, no obstante aún existen muchos lugares que por su ubicación geográfica no tienen acceso a Internet.

Otro avance que ha ayudado mucho a la disminución de la brecha digital por el factor geográfico es el aprendizaje virtual o a distancia “E-learning”, modalidad de estudio que permite compartir conocimiento mediante aulas virtuales y la autoeducación valiéndose de información publicada en el internet por estudiosos de todo el mundo.

### **3.2.6. La educación.**

El nivel de educación digital genera una desventaja tecnológica en la población. De acuerdo a lo publicado por (INEC, 2013). En Ecuador el 29.2% de la población inciden en el analfabetismo digital; para calificar a una persona dentro del nicho de analfabetismo digital se considera tres características conjuntas: que no tenga celular, que no haya utilizado una computadora en los últimos 12 meses y que no haya utilizado el internet en los últimos 12 meses.

Una oportunidad de disminuir el analfabetismo digital es la forma atractiva y fácil de interacción que nos ofrece la Web 2.0, pues dentro de esta existen las redes sociales que han invitado a personas de todas las edades a interactuar y comunicarse por medio del internet. (Francisco Amat, 2011). También nos invita a formar a estudiantes con espíritu crítico, y orientados a potenciar la competencia comunicativa tanto de lectura como de escritura de mensajes mediáticos, de fácil entendimiento e interpretación. También nos menciona que:

El gran valor de la web 2.0 es reducir dramáticamente la distancia entre los que acceden a la web y los que publican en ella información. Actualmente cualquier usuario puede acceder, de forma gratuita, a un gestor de contenidos en la forma de un blog, publicar imágenes o vídeo (Peña, Córcoles, & Casado, 2006). En este sentido, cabe destacar el abaratamiento de la producción y distribución de material escrito y audiovisual (Barolome & Willem, 2008). La web 2.0 ofrece nuevas funcionalidades que permiten hablar de Internet no sólo como gran fuente de recursos, sino, además, como la plataforma donde trabajar con esos recursos. Esto se traduce en dos acciones sustantivas del proceso de aprendizaje: generar contenidos y compartirlos (Peña et al., 2006).

Lo mencionado en el párrafo anterior nos indica las diversas bondades y facilidades que nos permite la Web 2.0, y que han apoyado a disminuir el analfabetismo digital y con ello la brecha digital existente por este factor.

### **3.2.7. La integridad física.**

Las personas con discapacidades físicas en su mayoría están limitadas a pasar en sus domicilios, lo cual les imposibilita relacionarse con otras personas en la sociedad. Con ello la tecnología les abre la oportunidad de relacionarse, opinar y ser críticos ante la

sociedad. En el informe presentado por la (UNESCO, 2005) mencionan que el internet representa para las personas con discapacidad física una posibilidad única de reinserción social, por ejemplo mediante el teletrabajo. Sin embargo, los discapacitados acumulan desventajas económicas, culturales o psicológicas que contribuyen a ahondar la brecha digital. No obstante, es preciso reconocer los esfuerzos de los constructores para crear instrumentos que facilitan la utilización de los ordenadores por los discapacitados, por ejemplo la posibilidad de acceso a menús contextuales utilizando el teclado con una sola mano (UNESCO, 2005).

Las redes sociales también aportan a la integración de personas con discapacidades a la sociedad de la información, ayudando con ello a la disminución de la brecha digital. En estudio emprendido por (Suría, 2012) respecto a “Redes sociales online y su utilización para mejorar las habilidades sociales en jóvenes con discapacidad” nos manifiesta que los jóvenes con discapacidad mantienen la misma dirección sobre el uso y preferencias hacia las redes sociales que tienen los jóvenes sin discapacidad, sin embargo, y aunque es evidente que estos espacios son utilizados por usuarios sin ningún tipo de discapacidad, para algunos jóvenes con características específicas, como vivir con movilidad reducida puede significar un recurso especialmente útil a la hora de aportarles la posibilidad de reducir sus limitaciones e incrementar las vías de interacción.

Las redes sociales han dado la posibilidad a los jóvenes con discapacidades de ser parte de la sociedad de la información, mismos que de acuerdo al estudio mencionado en el párrafo anterior lo han sabido aprovechar y deshacer la desventaja frente a los jóvenes sin discapacidad, logrando disminuir en un gran porcentaje la brecha digital provocada por este factor (Suría, 2012).

Ahora luego de haber analizado cada factor que influye en la brecha digital, se puede palpar claramente como la incidencia hacia la brecha digital puede presentarse en cada caso por más de un solo factor, por ejemplo una persona que viva en un lugar que por su ubicación geográfica no tenga acceso a internet, y que adicional a ello sea parte de la población de analfabetos tecnológicos, hace que esta persona incida en dos factores que deberá superar para dejar de ser parte de la brecha digital, y formar parte de la sociedad de la información.

### **3.3. Influencia del internet en el rendimiento académico universitario.**

Se conoce como rendimiento académico a los resultados alcanzados por cada estudiante que cursa estudios académicos, ya sean los mismos básicos, medios o de nivel superior. Comúnmente el rendimiento académico de los estudiantes se lo asocia directamente con la inteligencia de los mismos, sin embargo el rendimiento académico alcanzado es el resultado de muchos otros factores, (Tapia Ermel, 2012) nos comentan que en el rendimiento académico intervienen muchas otras variables externas al sujeto, como la calidad del maestro, el ambiente de clase, la familia, el programa educativo y variables psicológicas. En las internas, como la actitud hacia la asignatura, la inteligencia, la personalidad, el auto-concepto del estudiante y la motivación. En suma, el rendimiento académico del estudiante depende de su situación material y social de existencia, que debe ser tomado en cuenta en el momento de evaluar su nivel de aprendizaje. De acuerdo a estas aseveraciones el rendimiento académico verdadero es el resultado del sacrificio de uno mismo, éxito satisfactorio, compensación de la perseverancia, respuesta positiva al interés y consagración de uno.

Los rendimientos académicos también están ligados directamente a la motivación de las personas por aprender, (Arias, Cabanach, Núñez Pérez, & González-Pienda, 1998) llegaron a la conclusión que el proceso de aprendizaje aumenta cuando el estudiante se siente auto-competente.

Hoy en día la tecnología ha ingresado a ser participe respecto al rendimiento académico de los estudiantes. Como nos comenta (Ayala, 2007), las TIC's sirven como herramienta para mejorar el aprendizaje a los estudiantes, y pueden tener un impacto positivo en el rendimiento académico. El uso de las TIC's da el direccionamiento necesario para analizar si las mismas están sirviendo positivamente al rendimiento académicos del estudiante o en su defecto solo sirve como diversión. El uso del internet como parte de las TIC's en el desempeño académico de los estudiantes se encuentra en relación directa a la habilidad para el manejo del Internet, la frecuencia con la que se use y la disponibilidad de acceso a la misma.

La desigualdad digital también tiene relación con el rendimiento académico de los estudiantes, (Castaño-Muñoz, 2010) presenta algunos determinantes de la TIC's con relación a este tema y son: el acceso a internet que se ha comprobado que tiene un impacto positivo en el aprovechamiento educativo de los estudiantes, la alfabetización

digital que les ayuda a mejorar sus calificaciones debido a que el conocimiento de tecnologías como web 2.0 entre otras que proporcionan más posibilidades de sobresalir académicamente, el tiempo en la red que puede afectar positiva o negativamente al resultado académico de los estudiantes de acuerdo al uso que le dé al internet cada uno, y adopción de la tecnología que hace referencia a la finalidad de uso que se le da al internet.

De acuerdo a estudio realizado por (Duart, Gil, Pujol, & Castaño, 2008) han demostrado que no solo basta con utilizar el tiempo de conexión a la red con fines académicos para alcanzar mejoras en el resultado académico, sino también son necesarias algunas condiciones como: la existencia de interés por aprender, y la integración del uso del internet en la pedagogía de la educación superior.

La web 2.0 también cumple un papel fundamental como influencia del internet para un buen resultado académico, ya que esta tecnología ha permitido aspectos importantes dentro del campo académico como los son: auto-aprendizaje, educación virtual o a distancia, foros de intercambio de conocimiento, paginas de proporción y obtención de conocimiento (blogs), redes sociales, entre otros. Todos estos aspectos están disponibles dentro de la gama de opciones que tienen actualmente los estudiantes para aprender y ser sobresalientes en su desempeño educativo (Francisco Amat, 2011).

Haciendo un resumen de lo descrito en este punto, podemos darnos cuenta con facilidad que el internet tiene una incidencia directa en el resultado académico de los estudiantes, el mismo puede ser positivo o negativo de acuerdo al manejo de los múltiples elementos que se deben considerar respecto al internet y la variación de los mismos en la aplicación orientada hacia el mejoramiento del aprovechamiento académico de los estudiantes.

#### **3.4. Tecnologías como herramienta académica en la Universidad.**

En el modelo de Universidad actual, la tecnología es imprescindible dentro del aprendizaje y el desempeño académico, por lo que se convierte en una herramienta básica para estudiantes, docentes y personal administrativo que conforman las entidades educativas de nivel superior, lo que los obliga a formar parte de la sociedad del conocimiento.

La inclusión de la Tecnología dentro de la universidad se la puede apreciar y considerar desde los siguientes aspectos, sabiendo que la unión de estos hace que las TIC`s funcionen como un ente principal e íntegro dentro de la sociedad universitaria:

- ✓ Políticas, directrices y manuales de herramientas tecnológicas.

Dentro de la institución que usa herramientas tecnológicas es esencial que se implementen y se difundan las políticas, directrices y manuales de administración y uso de herramientas tecnológicas empleadas en la Universidad, con la finalidad de tener lineamientos claros respecto a la buena administración y buen uso de dichas herramientas en el día a día, impidiendo políticamente que los usuarios realicen prácticas irregulares que estén fuera de los lineamientos establecidos institucionalmente.

- ✓ Infraestructura tecnológica.

La Infraestructura tecnológica es indispensable y la base para todos los demás aspectos, pues en esta involucramos: la adquisición y mantenimiento de software educativo y administrativo, servidores, computadoras, equipos de Telecomunicaciones, Accesibilidad a Internet, biblioteca virtual, y en especial infraestructura que cada Universidad tenga para impartir enseñanza con tecnología de vanguardia.

- ✓ Desarrollo y capacitación continua.

En este aspecto hablaremos sobre los docentes, pues para que la tecnología tenga efecto positivo en la educación superior es necesario tener docentes que para sus cátedras utilicen herramientas tecnológicas y para aquellos que no las utilicen por desconocimiento tecnológico, la institución debería mantener un programa de capacitación continua para cubrir esta brecha en los docentes. Así también la institución debería programar capacitaciones a los estudiantes respecto a herramientas nuevas y de uso recurrente en su rol de estudiante.

- ✓ Programa de incentivos, respecto a utilización de tecnología.

Los programas de incentivos siempre han dado resultado para los docentes y estudiantes que son reacios a la utilización de la tecnología, ya sea por

desconocimiento o simplemente por su miedo al cambio. Ante ello la institución suele aplicar esta estrategia con la frecuencia que ellos lo creen conveniente respecto a estudios aceptación tecnológica de su población estudiantil y docente.

- ✓ La inserción de tecnología en el aula.

Las TIC`s en el aula permiten que docente y estudiantes puedan interactuar eficientemente y da dinamismo a la forma de aprendizaje aportando positivamente a la educación.

- ✓ Desarrollo de mallas de Investigación.

Si bien la universidad tradicional ya emprendía con sus docentes y estudiantes diferentes investigaciones, la mayoría de ellas se realizaban de forma local y limitada. Las TIC`s han permitido emprender en mallas de investigación mediante canales virtuales expandiendo las mismas a nivel internacional y con metas desafiantes y de gran impacto educativo, económico y social.

- ✓ Indicadores de seguimiento y resultados.

Los indicadores permiten evidenciar los resultados alcanzados en cada uno de los proyectos que se han emprendido, o temas que se quiere dar seguimiento. Las TIC`s permitieron dar una giro de 360 grados a las mismas, y en el ámbito educativo permiten el manejo de indicadores tanto a estudiantes, alumnos y altos directivos de la institución.

También las TIC`s han colaborado de manera importante para el desarrollo y evolución del sector educativo superior, algunos de sus principales logros que se han podido alcanzar por las TIC`s son:

- Educación a distancia.
- Bibliotecas virtuales.
- Apoyo a enseñanza en las aulas.
- Aula Virtual.
- Campus virtual (acceso a internet en el campus universitario).
- Mallas de investigación internacional.

- Reuniones virtuales.

De acuerdo a lo aportado por (Marín, 2011), estos logros se pueden consolidar en tres beneficios como los son: la innovación, la apertura del sistema al exterior y la posibilidad de desarrollar comunidades virtuales en aprendizaje entre docentes universitarios.

En estos últimos años, las instituciones de control a la educación superior en Ecuador frecuentemente están evaluando a las instituciones educativas respecto las TIC's como parte de la oferta académica a los estudiantes, y esto influye para otorgar una categoría de calidad a cada institución (CEAACES, 2012). Lo cual ha obligado a las universidades a crear proyectos de renovación tecnológica en su infraestructura e incentivar tanto a docentes como a estudiantes a utilizar las TIC's en su rol dentro de la institución educativa.

### **3.5. Aprendizaje Digital**

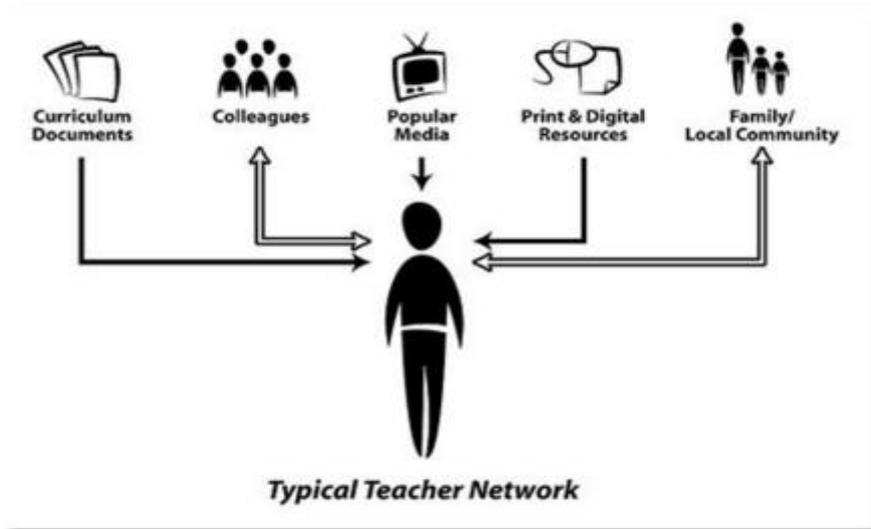
Con la fiebre tecnológica que vivimos hoy en día, el aprendizaje digital es un método de estudio habitual en todos los centros de estudio superior. En Ecuador la mayoría de Universidades ya cuenta con un campus virtual; también acceso a internet en todas sus instalaciones, y adicional han creados los patios de tareas donde los estudiantes utilizando sus equipos tecnológicos personales realizan sus trabajos extra-clase. También en las Universidades se han implementado las bibliotecas virtuales, logrando con ello que múltiples estudiantes tengan acceso simultáneo al mismo libro, lo que en una biblioteca tradicional no es posible. Adicionalmente se puede tomar como parte del aprendizaje digital la utilización de tecnologías dentro del aula por parte del profesor para aplicar sus cátedras.

(Adell, 2011) nos expresa una nueva forma de aprendizaje personalizado, orientado a la colaboración de usuarios participantes por medios virtuales, utilizando las bondades del internet y aplicaciones de vanguardia que en ella habitan, dicha forma de aprendizaje es llamada PLE (Entornos Personales de Aprendizaje), en donde manifiesta que los tiempos están cambiando muy deprisa y constantemente están surgiendo tecnologías, herramientas y servicios que abren un mundo de posibilidades de comunicación y aprendizaje que nos permiten acceder a la información y el conocimiento de una forma personalizada, adaptada a nuestras inquietudes,

conocimientos y necesidades, mejorando las posibilidades de acceso a contenidos, a recursos, a fuentes de información, a experiencias, a opiniones, a ideas, posibilitando la creación de una red de personas con la que compartimos inquietudes e intereses. Esta situación, de sobreabundancia de posibilidades, que nos puede parecer caótica o incluso inabordable no debería impedir que tomemos conciencia de cuáles son nuestros entornos personales de aprendizaje, también llamados PLE, para tratar de rediseñarlos, en caso de ser necesario, para mejorar el control y la gestión de nuestro propio aprendizaje. PLE es un conjunto de sistemas que ayudan a los estudiantes a tomar el control y gestión de su propio aprendizaje. Esto incluye el apoyo a los estudiantes para fijar sus propios objetivos de aprendizaje, gestionar su aprendizaje, la gestión de los contenidos y procesos, comunicarse con otros en el proceso de aprendizaje y lograr así los objetivos de aprendizaje. Un PLE puede estar compuesto de uno o varios subsistemas: así, puede tratarse de una aplicación de escritorio o bien estar compuestos por uno o más servicios web. Los PLE incluyen la integración de los episodios de aprendizaje formales e informales en una experiencia única, el uso de redes sociales que pueden cruzar las fronteras institucionales y la utilización de protocolos de red (*Peer-to-Peer*, servicios web, sindicación de contenidos) para conectar una serie de recursos y sistemas dentro de un espacio gestionado, [Ver Figuras 10A y 10B].

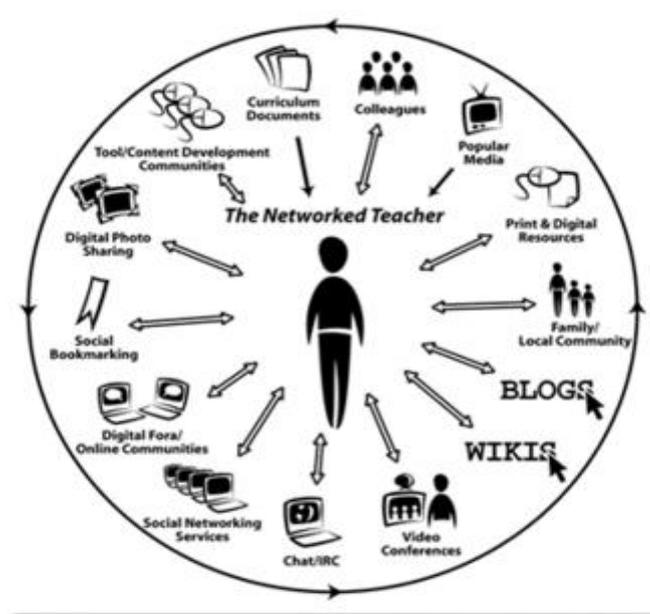
Adicionalmente una PLE tiene tres características (Adell, 2011):

- Cada alumno se fija sus propios objetivos de aprendizaje.
- No hay evaluaciones, ni títulos; no hay una estructura formal.
- Posibilidad que nos brinda Internet para disponer de un conjunto de herramientas y recursos gratuitos para compartir y aprender a través de ellos.



*Los docentes antes*

**Figura 10A.** Dos enfoques de docentes Universitarios (Antes).  
Fuente: (Adell, 2011).



*Los docentes ahora*

**Figura 10B.** Dos enfoques de docentes Universitarios (Ahora).  
Fuente: (Adell, 2011).

## **CAPÍTULO IV - MINERIA DE DATOS**

#### **4. Minería de Datos**

La minería de datos es un proceso que permite encontrar información útil y comprensible a partir de grandes bases de datos, utilizando para ello análisis matemático para deducir los patrones y tendencias que existen en los datos (MSDN Microsoft). También (Aluja, 2001) en su trabajo “La Minería de Datos, entre la Estadística y la Inteligencia Artificial” define a la minería de datos como una técnica emergente que parte, por un lado de las técnicas estadísticas y por otro de las técnicas de inteligencia artificial. Estas dos raíces principales que conforman la minería de datos le han permitido prevalecer en el tiempo solventando problemas de crecimiento de datos adheridos a los avances tecnológicos; la estadística se encarga del análisis de los datos (estudio de la población, variabilidad, generalidad), mientras que la inteligencia artificial se preocupa de establecer soluciones algorítmicas mediante sus técnicas específicas (patrones secuenciales, clustering, clasificación y predicción); ayudando a tomar decisiones inteligentes de una manera fácil sobre problemas complejos.

Para emprender un proyecto utilizando Minería de datos es necesario considerar como mínimas las siguientes etapas que se debe cumplir para una efectiva aplicación:

1. Identificación de oportunidad de mejora, en donde la minería de datos puede proporcionar valor.
2. Modelamiento de la solución, utilizando minería de datos.
3. Actuar sobre la información obtenida de la MD.
4. Medición de los resultados.

##### **4.1. Áreas o Mercados en lo que se aplica la minería de datos**

La minería de datos actualmente se puede aplicar en todas las áreas, puesto que el origen de la data para el análisis puede ser obtenida de sus sistemas informáticos o a través de encuestas, entrevistas, etc.; (Riquelme, Ruiz, & Gilbert, 2006) entre algunas de las áreas que más se ha potencializado en los últimos años la aplicación de la minería de datos tenemos:

#### **4.1.1. Banca y Comercio**

Es aplicado principalmente para: segmentación de clientes, análisis de riesgos, previsión de ventas, presupuestos, e indicadores de cumplimiento de metas, fidelidad de los clientes.

#### **4.1.2. Medicina y Farmacia**

En el campo de la medicina es muy útil para diagnóstico de enfermedades, y aprendizaje de sistemas expertos en base a historial de cada paciente, logrando con ello mayor efectividad en los tratamientos.

#### **4.1.3. Seguridad y Detección de Fraude**

La minería de datos en esta área ayuda a temas como: reconocimiento facial, acceso a redes no permitidas, alertas de phishing a través de técnica de transacciones usuales.

#### **4.1.4. Sistema Educativo**

En el sistema educativo que es nuestro foco de estudio, la minería de datos nos permite obtener resultados de: rendimiento académico de los estudiantes, eficiencia y eficacia en técnicas de enseñanza y aprendizaje, análisis de deserción de estudiantes, aprovechamiento de servicios prestados al estudiante como complemento a su enseñanza en el aula.

#### **4.1.5. Geología, Minería y Agricultura**

Permite la identificación de áreas de uso para distintos cultivos, explotación minera en bases de datos de imágenes de satélites, entre otras.

#### **4.1.6. Ciencias Ambientales**

En las ciencias ambientales la minería de datos principalmente es aplicada para la identificación de modelos de funcionamiento de ecosistemas naturales o artificiales.

#### **4.1.7. Ciencias Sociales**

La minería de datos en el área de las ciencias sociales permite realizar estudios de los flujos de opinión y comportamiento social de la población.

Es necesario mencionar que en el Ecuador, ninguna institución gubernamental o privada ha emprendido hasta el momento un análisis poblacional jurídico (empresarial), para evidenciar los niveles de aplicación de minería de datos en las diferentes áreas de acuerdo a la actividad económica a la que pertenece cada una respectivamente.

#### **4.2. Tipos de Bases de Datos**

La minería de datos se puede aplicar a todo tipo de bases de datos imaginable, esto desde numéricos hasta imágenes satelitales. Con ello podríamos decir que “cualquier cosa” constituye un dato como tal.

Los datos se pueden agrupar principalmente de la siguiente manera respecto a su tipo:

- Bases de datos relacionales. (variables, columnas, campos)
- Bases de datos espaciales. (datos geográficos, imágenes médicas, redes de transporte).
- Bases de datos documentales. (principalmente documentos de texto)
- Bases de datos multimedia. (imágenes, audio, video)
- Bases de datos Word Wide Web, (todos los datos relacionados con la WWW, linked data).

Para este trabajo de investigación, se utilizara los datos agrupados en “Bases de datos relacionales” de acuerdo a lo expuesto en líneas anteriores, ya que la información que se tiene como input está dada por variables, y el objetivo es relacionar las diferentes variables de interés de acuerdo a los datos que cada una de estas contiene.

La Minería de datos es el proceso que tiene todos los componentes que se necesita para esta investigación, pues a diferencia de otros procesos de Bases de Datos, esta nos permite: descubrir información que no es esperada obtener, integrar diferentes tipos de datos, ofrece modelos confiables y se pueden construir de manera rápida (Riquelme et al., 2006), es por ello que se optó por aplicar el proceso de minería de datos

para llevar a cabo el análisis de información levantada para este trabajo de investigación.

### **4.3. Modelos y Tareas de Minería de Datos**

Los modelos de Minería de datos permiten la transformación de datos en conocimiento. Se dice que los modelos son el producto de la minería de datos. Principalmente existen dos tipos de modelos de minería de datos, los descriptivos y los predictivos; y con esto modelos se puede realizar distintos tipos de tareas.

#### **4.3.1. Modelos Descriptivos y sus Tareas**

Tienen como objetivo la identificación de patrones que expliquen o describan los datos ya existentes, en estos modelos se pueden realizar las siguientes tareas:

##### **4.3.1.1. Clustering o agrupamiento**

Su objetivo es la agrupación o segmentación de datos homogéneos extraídos de las clases de un dominio establecido. Por ejemplo este modelo puede aplicarse para lograr segmentar los clientes de un Banco según sus perfiles comerciales; o agrupar los usuarios de internet de acuerdo al propósito para el que la utilizan. Existen dos tipos de clusterización:

- Jerárquico.\_ Cuando los datos se agrupan de manera arborescente.
- No Jerárquico.\_ Cuando generan particiones a nivel paramétrico (asumen cierta forma paramétrica en el modo de agrupamiento de los objetos) y no paramétricos (no asumen nada sobre el modo de agrupamiento de los objetos).

##### **4.3.1.2. Correlación**

El modelo de correlación tiene como objetivo poder ver, si teniendo dos variables (A y B) existe correlación entre estas, tomando en cuenta que la correlación se muestra si al aumentar los valores de A lo hacen también los de B y viceversa. Por ejemplo se puede aplicar este modelo para analizar la correlación de utilización de herramientas IT por los usuarios.

#### **4.3.1.3. Asociación**

Expresan patrones de comportamiento en los datos. Este modelo permite describir las relaciones o asociaciones que existen entre los valores de las variables de un dominio establecido. Por ejemplo este método se puede aplicar en el análisis de productos de un almacén que los clientes llevan en conjunto, con la finalidad de reorganizarlos dentro del almacén.

#### **4.3.2. Modelos Predictivos y sus Tareas**

Tienen como objetivo la estimación de valores de variables de interés a predecir, esto a partir de otras variables que cumplen el papel de predictoras, en estos modelos se pueden realizar las siguientes tareas.

##### **4.3.2.1. Regresión**

En donde la variable a predecir es continua, este modelo nos ayuda a encontrar la similitud entre varios valores de atributos de una clase establecida. Por ejemplo este modelo permite predecir y determinar el número de stock que se necesita en un almacén comercial respecto a cada producto para poder atender a todos los clientes partiendo de un análisis de ventas históricas.

##### **4.3.2.2. Clasificación**

En donde la variable a predecir es discreta (nominal u ordinal), este modelo tiene como objetivo poder clasificar un dato dentro de una o varias clases establecidas. Por ejemplo este modelo permite clasificar los clientes rentables de un almacén de acuerdo a la frecuencia de sus compras.

#### **4.4. Técnicas o Métodos de Minería de Datos**

Nos permiten obtener un modelo de conocimiento, mismo que presenta los patrones de comportamiento observados en valores de las variables en cuestión o relaciones de asociación entre dichas variables. A continuación se describen las técnicas más utilizadas.

#### **4.4.1. Métodos Bayesianos**

Una de las características primordiales de los métodos bayesianos es el uso de distribuciones de probabilidad para cuantificar incertidumbre de los datos que se desea modelar. Estos métodos proporcionan una metodología práctica para la inferencia y predicción y en última instancia para tomar decisiones que involucran cantidades inciertas (Hernández Orallo, Ramírez Quintana, & Ferri Ramírez, 2004).

Esta técnica es muy utilizada para problemas de inteligencia artificial, utilizando para ello el teorema de Bayes. La ventaja que nos ofrece esta técnica es que nos permite estimar probabilidades en grandes cantidades de datos. Y la desventaja es que no se puede realizar predicciones en pocos datos, ya que al no tener la suficiente cantidad de información proporcionada, este presentaría un modelo incorrecto.

#### **4.4.2. Reglas de Asociación y Dependencia.**

Esta técnica radica en que mediante reglas se expresan patrones de comportamiento entre los datos de las clases del dominio en función de la aparición conjunta de valores de dos o más atributos (Hernández Orallo et al., 2004).

Una de las aplicación más típicas de esta técnica ha sido en los sistemas de comercio electrónico para informar sobre las preferencias de compra de los clientes. Su aplicación a sistemas de e-learning permite descubrir relaciones o asociaciones entre distintas páginas Web visitadas.(C. Morales, Soto, & Martínez, 2005). Para extraer los patrones de comportamiento esta técnica utiliza algoritmos A priori.

#### **4.4.3. Métodos relacionales y estructurales.**

La mayoría de las técnicas de minería de datos trabajan con los datos en el formato atributo-valor, pues tales técnicas asumen que los datos vienen en una única tabla. En la presente técnica se utilizan un lenguaje de representación relacional mucho más ponente que los lenguajes tradicionales de presentación, pues se analiza los datos de las tablas relacionadas. (Hernández Orallo et al., 2004).

Este método nos ayuda a descubrir patrones de comportamiento más complejos, haciendo uso de la relación existente entre las tablas relacionadas, sin tener necesidad de tener los datos en una sola tabla.

Una desventaja que presenta esta técnica es que es mucho menos eficiente respecto a la velocidad de respuesta que nos brindan otras técnicas.

#### **4.4.4. Métodos basados en casos y en vecindad.**

Estos métodos resuelven problemas partiendo de información extraída de un conjunto de ejemplos ya existentes previamente para poder compararlos con los nuevos casos. Respecto a métodos basados en vecindad, su predicción se basa fundamentalmente en la utilización del conjunto de ejemplos “vecinos” al dato que hay que procesar.(Hernández Orallo et al., 2004).

Para esta técnica se utiliza algoritmos: jerárquicos (Two-step, COBWED), vecinos más próximos (K-NN), no jerárquicos (K-means).

#### **4.4.5. Árboles de decisión.**

Respecto a esta técnica (Hernández Orallo et al., 2004) aporta que *“la técnica basada en árboles de decisión es quizá el método más fácil de utilizar y de entender”*. El árbol de decisión presenta una arquitectura jerárquica, la misma que está formada por nodos, y cada uno de estos nodos permite establecer una regla o condición, estas reglas pueden tener valores de respuesta verdaderos o falsos, de tal forma que el resultado final sea el cumplimiento de todas las condiciones desde el nodo raíz hasta alguno de sus nodos hijos.

Los arboles de decisión permiten segmentar la población de datos para encontrar grupos homogéneos respecto a una cierta variable de respuesta. Los modelos que se apegan a esta técnica son: la clasificación, regresión y agrupamiento.

La ventaja más significativa que nos brinda esta técnica es que las opciones de resultado posibles que nos presentan los arboles de decisiones son precisas, lo que nos ayuda a alcanzar una sola decisión a tomar.

#### **4.4.6. Redes Neuronales.**

Esta técnica es basada en el funcionamiento de sistema nervioso humano, mismos que aprenden tareas o modelos mediante el entrenamiento de los pesos (valores).

Tiene una arquitectura secuencial, el procesamiento de la información pasa por elementos simples llamados neuronas. Las redes neuronales permiten procesamiento paralelo, estas pueden ser adaptativas y asociativas.

Una red neuronal trabaja con un conjunto de unidades de procesamiento, con un estado de activación de neuronas y con conexiones entre las neuronas.

(Hernández Orallo et al., 2004) indica que esta técnica posee dos tipos de aprendizaje: uno que es supervisado (se proporciona un conjunto de datos de entrada y la respuesta correcta), es útil en modelos de regresión y clasificación; y el aprendizaje no supervisado (se proporciona a la red un conjunto de datos de entrada), y la red debe auto-enseñarse para que pueda proporcionar la respuesta, este aprendizaje es útil para modelos de agrupamiento.

Una de las ventajas más significativas que nos presenta esta técnica es la capacidad que tiene de trabajar con información incompleta o inconsistente, ya que existen casos que se desea modelar con este tipo de características. Y una desventaja es que los modelos aprendidos son de difícil comprensión.

#### **4.4.7. Análisis Factorial (factorización).**

El análisis factorial es una técnica estadística que permite la reducción de datos, mediante la correlación de un conjunto de variables dadas, obteniendo como resultado un menor número de variables nuevas llamadas factores.

El análisis factorial puede ser de tipo exploratorio o confirmativo. Se considera exploratorio cuando no se conocen los factores correlacionados y por tanto se determinan como parte del resultado del mismo análisis; en cambio el análisis factorial confirmativo se aplica cuando ya se conoce los factores con los que se quiere trabajar respecto a la correlación de las variables, y se confirma dicha correlación de los factores a través de los análisis de comprobación o prueba de relación.

#### **4.4.8. Modelización estadística paramétrica.**

La modelación estadística paramétrica tiene como objetivo explicar el comportamiento de una variable a partir de otras; es una formalización de relaciones entre variables representadas a través de ecuaciones.

Por ejemplo, el saldo total bancario de clientes de una determinada edad, con un mismo nivel profesional y residiendo en una misma ciudad no es igual para todos ellos sino que sigue una cierta distribución; pero si sabemos la edad de la persona, su nivel profesional y su ciudad de residencia podemos dar una aproximación de su saldo bancario. En este ejemplo las variables a analizar son: saldo bancario como variable de salida (output), conocida también como de respuesta, mientras que la variable edad, nivel profesional y ciudad de residencia se consideran como de entrada (input) (Hernández Orallo et al., 2004).

Esta técnica se utiliza principalmente sobre modelos de minería de datos predictivos que son la regresión y la clasificación.

#### **4.4.9. Modelización estadística no paramétrica.**

Esta técnica permite construir modelos más flexibles, esto en comparación con la estadística paramétrica, ya que son capaces de modelar fenómenos complejos que involucran el análisis de un gran volumen de datos (Hernández Orallo et al., 2004).

Este método es recomendado cuando no se puede asumir que los datos se ajustan a una distribución conocida; y se utiliza principalmente sobre modelos de minería de datos predictivos que son la regresión y la clasificación

### **4.5. Algoritmos de Minería de Datos**

#### **4.5.1. Algoritmos evolutivos y reglas difusas.**

A través de los algoritmos Evolutivos se puede utilizar un procedimiento determinístico para alcanzar una solución óptima, comienza desde un punto aleatorio y se basa en una regla de transición especificada previamente para determinar la dirección de la búsqueda (Hernández Orallo et al., 2004).

Las principales ventajas del uso de los Algoritmos Evolutivos en el descubrimiento concreto de reglas de predicción es su capacidad para realizar una búsqueda global y el tratamiento óptimo del problema de la interacción de los atributos que la mayoría de los algoritmos voraces de inducción de reglas que se han utilizado tradicionalmente (Romero, Ventura, & Castro, 2003).

Un algoritmo evolutivo se basa en mantener una población de posibles soluciones del problema a resolver, llevar a cabo una serie de alteraciones sobre las mismas y efectuar una selección para determinar qué soluciones permanecen en generaciones futuras y cuales son eliminadas. Existen cuatro tipos de algoritmos evolutivos bien definidos que son los algoritmos genéticos, las estrategias de evolución, la programación evolutiva y la programación genética (González, 2007).

La lógica difusa trata con conjuntos difusos y conectores lógicos para modelar los problemas de razonamiento del mundo real de la misma forma que los hacen los seres humanos, modelando la vaguedad y manejando incertidumbre (González, 2007).

#### **4.5.2. Algoritmos de minería de datos en función del objetivo**

Los objetivos de la minería de datos se definen por el uso que se pretende del sistema; tenemos dos tipos de objetivos: la verificación en donde el sistema se limita a verificar la hipótesis del usuario; y el descubrimiento en donde el sistema encuentra nuevos patrones de forma autónoma (González, 2007), este último se puede descomponer en:

- **Supervisados o Predictivos.**\_ En estos el sistema encuentra patrones para predecir el comportamiento futuro, utiliza un conjunto de variables de la base de datos para predecir valores futuros desconocidos de otras variables de interés (González, 2007).
- **No supervisados o Descriptivos.**\_ En estos el sistema encuentra patrones para presentarlos a un experto en una forma comprensible para él, describen y aportan información de interés sobre el problema y el modelo que subyace bajo los datos (González, 2007).

#### **4.5.3. Algoritmos de clustering o agrupamiento.**

Estos algoritmos nos permiten crear grupos de datos con características similares u homogéneas. Estos algoritmos se dividen en tres grupos de acuerdo al método de agrupamiento que se utilice, es decir que tenemos los algoritmos de métodos jerárquicos, algoritmos de métodos particionales y los algoritmos de métodos basados en densidad (Pascual, Pla, & Sánchez, 2007).

#### 4.5.3.1. Algoritmos jerárquicos

Son aquellos en los que se va particionando el conjunto de datos por niveles, de modo tal que en cada nivel generalmente , se unen o se dividen dos grupos del nivel anterior, según si es un algoritmo aglomerativo o divisivo (Pascual et al., 2007). Los principales algoritmos que pertenecen a este grupo son:

- ✓ **Algoritmo Chameleon.**\_ Se aplica en dos fases; en la primera construye el grafo de los k vecinos más cercanos y usa un algoritmo de particionamiento de grafo para agrupar los puntos en subgrupos. Y en la segunda fase, usa un algoritmo jerárquico aglomerativo para encontrar los clusters genuinos combinando repetidamente estos subgrupos. En esta segunda fase determina el par de subgrupos más similares tomando en cuenta su interconectividad y cercanía, éstas expresan las características internas de los subgrupos, el modelo no es estático si no que es capaz de adaptarse a las características internas de los subgrupos según estos van cambiando (Pascual et al., 2007).
  
- ✓ **Algoritmo de criterio de aislamiento de grupos.**\_ Se integra un criterio de aislamiento de grupos en un algoritmo de agrupamiento jerárquico aglomerativo. Define el incremento de la disimilaridad o gap entre dos grupos  $C_i$  y  $C_j$ , el criterio de aislamiento es el siguiente: dados dos grupos  $C_i$  y  $C_j$  candidatos para unir, si  $gap_i \geq t_i$  ( $gap_j \geq t_j$ ) aislar el cluster  $C_i$  ( $C_j$ ) y continuar la estrategia de agrupamiento con el resto de los patrones. Si ninguno de los grupos excede el límite del gap unirlos. Emplea un umbral dinámico ( $t_i$ ) que va variando a lo largo del algoritmo y en dependencia del grupo  $C_i$  (Pascual et al., 2007).
  
- ✓ **Algoritmo COBWEB.**\_ Se caracteriza porque utiliza aprendizaje incremental, es decir que realiza las agrupaciones instancia a instancia. Durante la ejecución del algoritmo se forma un árbol (árbol de clasificación) donde las hojas representan los segmentos y el nodo raíz engloba por completo el conjunto de datos de entrada (Rodríguez, Cuadrado, & Sicilia, 2007).

#### 4.5.3.2. Algoritmos particionales

Son los que realizan una división inicial de los datos en grupos y luego mueven los objetos de un grupo a otro según se optimice alguna función objetivo. Estos algoritmos asumen un conocimiento a priori del número de clusters en que debe ser dividido el conjunto de datos, llegan a una división en clases que optimiza un criterio predefinido o función objetivo (Pascual et al., 2007). Los principales algoritmos que pertenecen a este grupo son:

- ✓ **Algoritmo K-means.** Es uno de los más simples y conocidos algoritmos de agrupamiento, sigue una forma fácil y simple para dividir una base de datos dada en  $k$  grupos (fijados a priori). La idea principal es definir  $k$  centroides (uno para cada grupo) y luego tomar cada punto de la base de datos y situarlo en la clase de su centroide más cercano, para luego recalculer el centroide de cada grupo y volver a distribuir todos los objetos según el centroide más cercano; este proceso se repite hasta que ya no hay cambio en los grupos (Pascual et al., 2007).

El nombre K-means le viene porque representa cada uno de los clusters por la media (o media ponderada) de sus puntos, es decir, por su centroide (Rodríguez et al., 2007).

El algoritmo K-means utiliza una medida de similaridad basada en el error cuadrático, esto con la finalidad que el algoritmo pueda generar mejores resultados [ver Figura 11].

$$E = \sum_{i=1}^k \sum_{p \in C_i} |p - m_i|^2$$

**Figura 11.** Error cuadrático K-means.

Fuente: (Instituto Nacional de Astrofísica, 2012).

El problema de la aplicación de este algoritmo es que falla cuando los puntos de un grupo están muy cerca del centroide de otro grupo, o cuando los grupos tienen diferentes tamaños y formas (Pascual et al., 2007). Adicional este

algoritmo únicamente se puede aplicar a atributos numéricos (Rodríguez et al., 2007).

- ✓ **Algoritmo CURE.**\_ Es un algoritmo híbrido entre los dos enfoques jerárquico y particional, mismo que trata de emplear las ventajas de ambos y de eliminar las limitaciones. En éste algoritmo en lugar de usar un solo punto como representante de un grupo se emplea un número **c** de puntos representativos del grupo. La similitud entre dos grupos se mide por la similitud del par de puntos representativos más cercanos, uno de cada grupo. Este algoritmo nos permite encontrar grupos de diferentes tamaños y formas (Pascual et al., 2007).
- ✓ **Algoritmo EM.**\_ El algoritmo EM (Expectation Maximization) inicia adivinando los parámetros de las distribuciones y los mismos utiliza para calcular las probabilidades de que cada objeto pertenezca a un cluster y usa esas probabilidades para re-estimar los parámetros de las probabilidades, hasta converger. Se usa en estadística para encontrar estimadores de máxima verosimilitud de parámetros en modelos probabilísticos que dependen de variables no observables.

#### 4.5.3.3. Algoritmos basados en densidad

Estos algoritmos enfocan el problema de la división de una base de datos en grupos teniendo en cuenta la distribución de densidad de los puntos, de modo tal que los grupos que se forman tienen una alta densidad de puntos en su interior mientras que entre ellos aparecen zonas de baja densidad (Pascual et al., 2007). Los principales algoritmos que pertenecen a este grupo son:

- ✓ **Algoritmo DBSCAN.**\_ Es el primer algoritmo basado en densidad, se definen los conceptos de punto central (puntos que tienen en su vecindad una cantidad de puntos mayor o igual que un umbral especificado), borde y ruido. Este algoritmo comienza seleccionando un punto **p** arbitrario, si **p** es un punto central, se comienza a construir un grupo y se ubican en su grupo todos los objetos denso-alcanzables desde **p**. Si **p** no es un punto central se visita otro objeto del conjunto de datos. El proceso continúa hasta que todos los objetos han sido procesados. Los puntos que quedan fuera de los grupos formados se

llaman puntos ruido, los puntos que no son ni ruido ni centrales se llaman puntos borde. Un grupo puede tener más de un punto central (Pascual et al., 2007).

- ✓ **Algoritmo OPTICS.**\_ La motivación para la realización de este algoritmo se basa en la necesidad de introducir parámetros de entrada en casi todos los algoritmos de agrupamiento existentes que en la mayoría de los casos son difíciles de determinar, además en conjuntos de datos reales no existe una manera de determinar estos parámetros globales, este algoritmo trata de resolver este problema basándose en el esquema del algoritmo DBSCAN creando un ordenamiento de la base de datos para representar la estructura del agrupamiento basada en densidad, además puede hacer una representación gráfica del agrupamiento incluso para conjuntos de datos grandes (Pascual et al., 2007).
  
- ✓ **Algoritmos KNNCLUST.**\_ Propone utilizar como regla basada en densidad la de los  $k$  vecinos más cercanos ( $k$ -NN), para tratar bases de datos de alta dimensionalidad como imágenes de satélites. La regla  $k$ -NN ha sido extensamente usada en muchos métodos de clasificación donde existe un conjunto de objetos etiquetados. En este algoritmo se determina el número de grupos de manera automática. Necesita la entrada de un solo parámetro: la cantidad de vecinos (Pascual et al., 2007).
  
- ✓ **Algoritmo SNN.**\_ Este algoritmo primero encuentra los vecinos más cercanos de cada punto de la base de datos y define la similaridad entre cada par de puntos en términos de cuántos vecinos más cercanos los dos puntos comparten. Usando esta definición de similaridad elimina ruido y outliers, identifica puntos centrales y construye grupos alrededor de éstos, el uso de una definición de similaridad con los vecinos más cercanos compartidos elimina problemas de diferentes densidades mientras que con el uso de los puntos centrales maneja los problemas de forma y tamaño (Pascual et al., 2007).
  
- ✓ **Algoritmo DENCLUE.**\_ Este algoritmo se aplica en dos fases, dados  $N$  puntos de una base de datos  $D$  provista de una distancia  $d$ , y  $x \in D$ , define función de

densidad, gradiente y punto atractor. En la primera fase divide el hiperrectángulo del conjunto de datos en hipercubos de aristas de longitud  $2\sigma$ , y determina cuales son los hipercubos más poblados y los hipercubos que están conectados. En la segunda fase considera solamente los hipercubos más poblados y los conectados a hipercubos más poblados para determinar los grupos, para cada  $x$  en estos hipercubos determina el valor de la función de densidad pero considerando solamente aquellos puntos  $x'$  tales que su distancia al centro del hipercubo al que  $x$  pertenece sea menor o igual que  $4\sigma$ , similarmente halla el gradiente y el atractor de densidad para  $x$  para el que la función de densidad sea mayor o igual que un valor  $\xi$  y clasifica a  $x$  en la clase de su atractor. Con este algoritmo se obtiene el número de grupos de manera natural (Pascual et al., 2007).

#### **4.5.4. Algoritmos de clasificación.**

Estos Algoritmos tienen como objetivo poder clasificar un dato dentro de una o varias clases establecidas.

##### **4.5.4.1. Algoritmo de Naive Bayes**

Este algoritmo se basa en la suposición que las cantidades de interés se rigen por distribuciones de probabilidad, y que la decisión óptima puede tomarse por medio de razonar acerca de esas probabilidades junto con los datos observados. En tareas como la clasificación de textos este algoritmo se encuentra entre los más utilizados (VALERO, 2005).

##### **4.5.4.2. Algoritmo C4.5**

El esquema del algoritmo C4.5 fue diseñado como una extensión del algoritmo ID3, este último forma parte de los clasificadores conocidos como árboles de decisión, los cuales son árboles donde sus nodos internos son etiquetados como atributos, las ramas salientes de cada nodo representan pruebas para los valores del atributo, y las hojas del árbol identifican a las categorías. Estos algoritmos proporcionan un método práctico para aproximar conceptos y funciones con valores discretos (VALERO, 2005). Los pasos a seguir en C4.5 son:

1. Separar  $\Omega$  en conjunto de entrenamiento y conjunto de validación.

2. Construir el árbol de decisión para el conjunto de entrenamiento (aplicar ID3).
3. Convertir el árbol en un conjunto de reglas equivalente, donde el número de reglas es igual al número de posibles rutas desde la raíz a los nodos hoja.
4. Podar cada regla eliminando precondiciones que resulten en mejorar la exactitud en el conjunto de validación.
5. Ordenar las reglas descendientemente de acuerdo a su exactitud, y usarlas en ese orden para clasificar futuros ejemplos (VALERO, 2005).

#### **4.5.5. Algoritmos de asociación.**

Estos algoritmos permiten encontrar patrones de comportamiento entre los datos, en función directa de la aparición conjunta de valores de dos o más atributos, seguidamente se detalla uno de los más utilizados que es el Algoritmo Apriori.

##### **4.5.5.1. Algoritmo Apriori**

Este algoritmo inicia generando todos los itemsets con un elemento, después usa estos para generar los de dos elementos y así sucesivamente, luego se toma todos los itemsets que cumplan con el mínimo soporte (esto permite eliminar posibles combinaciones). Luego de haber cumplido todo esto, genera las reglas revisando que cumplan con el criterio mínimo de confianza (E. Morales & Escalante, 2015).

#### **4.5.6. Otros Algoritmos.**

Estos algoritmos principalmente permiten realizar factorial en atributos y análisis de comprobación de confiabilidad de los métodos aplicados.

##### **4.5.6.1. Algoritmo CfsSubsetEval**

Este algoritmo evalúa un subconjunto de atributos considerando la habilidad predictiva individual de cada variable, así como el grado de redundancia entre ellas. Se prefieren los subconjuntos de atributos que estén altamente correlacionados con la clase y tengan baja intercorrelación (E. Morales & Escalante, 2015).

#### **4.5.6.2. Algoritmo ChiSquaredAttributeEval**

Este algoritmo calcula el valor estadístico Chi-cuadrado de cada atributo con respecto a la clase y así obtiene el nivel de correlación entre la clase y cada atributo (E. Morales & Escalante, 2015).

### **4.6. Herramientas de Minería de Datos**

Las herramientas facilitan el desarrollo de los modelos y tareas que utilizamos para la extracción de conocimiento, las mismas que tiene programados los algoritmos específicos para la aplicación de técnicas de minería de datos, adicional estas herramientas en su mayoría presentan una interfaz gráfica y de fácil comprensión en el manejo para los usuarios. Las más utilizadas son:

#### **4.6.1. Weka**

Weka (*Waikato Environment for Knowledge Analysis*), es una herramienta tecnológica grafica de libre distribución bajo licencia GNU-GPL; fue desarrollada por los investigadores de la Universidad de Waikato en Nueva Zelanda.

Esta herramienta permite aplicar técnicas de aprendizaje automático (como redes neuronales, arboles de decisión, clustering y reglas de asociación) y de minería de datos en general; permite el acceso a datos a través de un archivo en formato ARFF(archivo plano) y posee una interfaz gráfica para el usuario.

Las ventajas de esta herramienta es que al estar completamente programada en Java permite correr la misma en la mayoría de las plataformas, contiene muchos algoritmos de técnicas de minería de datos para poder aplicar los diferentes modelos.

#### **4.6.2. Spss Clementine**

Es una herramienta tecnológica estadística muy usado por las empresas de investigación de mercado, no es de distribución libre y actualmente pertenece a IBM. Permite trabajar con gran cantidad de datos y aplicar técnicas de minería de datos como redes neuronales, reglas de asociación, adicional permite visualizar histogramas y diagramas de dispersión.

#### **4.6.3. ODMS (Oracle Data Mining suite)**

Esta es una herramienta comercial que pertenece a empresa Oracle, permite grandes volúmenes de información y su arquitectura es cliente / servidor.

Esta herramienta tiene acceso a datos mediante diversos formatos (Almacenes de datos, SQL, Oracle), contiene preprocesador de datos (muestreo de datos, patrones de datos), y tiene algoritmos de técnicas de aprendizaje (redes neuronales, regresión lineal). Adicional esta herramienta permite importar datos directamente en Excel.

#### **4.6.4. Kepler**

Kepler es una herramienta comercial que pertenece a la empresa Dialogis. Permite aplicar técnicas de minería de datos como redes neuronales, regresión no lineal y aplicaciones estadísticas. Esta herramienta posee interfaz gráfica para el usuario.

#### **4.6.5. RapidMiner**

Es una herramienta tecnológica gráfica de libre distribución bajo licencia AGPL; fue desarrollada por el departamento de inteligencia artificial de la Universidad de Dortmund.

Esta herramienta permite aplicar técnicas de aprendizaje automático (como redes neuronales, arboles de decisión y reglas de asociación) y de minería de datos en general; permite la utilización de algoritmos incluidos en Weka. Su página oficial es <https://rapidminer.com/>

#### **4.6.6. Db2 Intelligent Miner**

Herramienta tecnológica gráfica comercializada por IBM, misma que trabaja bajo arquitectura cliente / Servidor y permite procesar grandes cantidades de datos, trabaja a través de los módulos: DB2 intelligent miner for data, DB2 intelligent miner scoring, DB2 intelligent miner modelling, DB2 intelligent miner visualization.

#### **4.6.7. Statistica Data Miner**

Es una herramienta tecnológica desarrollada y comercializada por la empresa StatSoft, permite grandes volúmenes de información.

Esta herramienta contiene preprocesador de datos (muestreo de datos, patrones de datos), y tiene algoritmos de técnicas de aprendizaje (redes neuronales, regresión lineal, regresión múltiple). Adicional esta herramienta permite importar datos directamente en Excel, Oracle, SQL. Su página oficial es <http://www.statsoft.com/>

#### **4.6.8. Orange**

Orange es una herramienta Tecnológica que permite realizar minería de datos; fue desarrollado en la facultad de informática de la Universidad de Liubliana en Eslovenia. Tiene componentes desarrollados en lenguaje C++ que permiten implementar algoritmos de minería de datos y presentaciones graficas de datos. Posee un entorno gráfico para el usuario y es distribuido bajo licencia GPL. Su página oficial es <http://orange.biolab.si/>

### **4.7. Procesos para la aplicación de la Minería de Datos**

De acuerdo al análisis emprendido por (Azevedo & Santos, 2008) los procesos que se puede utilizar para la aplicación de la minería de datos son KDD, SEMMA y CRISP-DM.

#### **4.7.1. Proceso KDD**

Uno de los procesos utilizados para aplicar la minería de datos es la KDD (Knowledge discovery in databases). Dicho proceso debe cumplir las siguientes fases mínimas para poder alcanzar una solución óptima.

##### **4.7.1.1. Recopilación e integración de datos.**

Esta fase debe determinar las fuentes de información que pueden ser útiles, como conseguir las y el diseño del esquema del almacén de datos que nos permita unificar de manera operativa toda la información recogida.

#### **4.7.1.2. Selección, Limpieza y transformación de datos**

Se debe seleccionar la información (variables) que se va a utilizar para el trabajo de minería de datos; adicionalmente es necesario eliminar el mayor número posible de datos con error para lograr calidad de la información a utilizar. Dentro de esta fase también se debe realizar de ser necesaria: la transformación de variables o registros de datos, e inclusive se puede realizar la construcción de nuevas variables que faciliten el proceso de minería de datos.

#### **4.7.1.3. Minería de datos.**

La minería de datos es la fase más característica del proceso de KDD (Knowledge Discovery in Databases). (Pereira, 2009) mediante las técnicas de minería de datos, se logra obtener modelos de conocimiento, que a su vez representan patrones de comportamiento observados en los valores de las variables del problema.

(Romero et al., 2005). Nos comparte que la aplicación de minería de datos en educación se puede ver desde dos orientaciones distintas:

- Hacia los autores.\_ Se aplican con la finalidad de obtener una mayor realimentación de la enseñanza de los profesores o autores hacia los alumnos, conocer más sobre como los estudiantes aprenden en la Web, evaluar a estudiantes por sus patrones de navegación, clasificarlos en grupos, etc.
- Hacia los estudiantes.\_ Se aplican con la finalidad de sugerir buenas experiencias de aprendizaje a los estudiantes, adaptación del curso según el progreso de aprendizaje, ayudar con sugerencias y atajos, entre otros.

Dentro de la fase de minería de datos existen algunos subprocesos a considerar, como lo son el reconocimiento de patrones, modelos de minería de datos y posibles técnicas de minería de datos a utilizar.

El reconocimiento de patrones o lectura de patrones es la ciencia que se encarga de la clasificación de objetos, personas, señales, representaciones, etc. Misma que trabaja en base a un conjunto previamente establecido de los posibles patrones a reconocer.

Respecto a los modelos y técnicas de minerías de datos, han sido explicadas a detalle en los puntos 4.3 y 4.4 de este documento.

#### **4.7.1.4. Evaluación e interpretación.**

Luego de haber obtenido el modelo, es necesario proceder a su validación, comprobando que las conclusiones que arroja sean válidas y suficientemente satisfactorias. De haber obtenido algunos modelos mediante el uso de distintas técnicas, se deben comparar los modelos con la finalidad de buscar aquel que se ajuste mejor al problema.

Si ninguno de los modelos obtenidos alcanza los resultados esperados, es necesario alterar alguno de los pasos anteriores para generar nuevos modelos.

#### **4.7.1.5. Difusión y Uso.**

Esta fase de difusión y uso contempla:

- Elaboración de informes para su distribución.
- Independencia en la forma de uso del nuevo conocimiento adquirido.
- Incorporarlo a sistemas ya existentes asegurándose de realizar las verificaciones respectivas para evitar inconsistencias y posibles conflictos.

La monitorización de la ejecución del sistema puede dar lugar a nuevos casos que realimentaran el ciclo del KDD.

#### **4.7.2. Proceso SEMMA**

El proceso SEMMA fue desarrollado por el Instituto SAS. El acrónimo SEMMA significa (Sample, Explore, Modify, Model, Assess) Para este proceso se considera un mínimo de cinco etapas para alcanzar una solución óptima (Azevedo & Santos, 2008).

##### **4.7.2.1. Muestreo**

Consiste en el muestreo de los datos mediante la extracción de una porción de un gran conjunto de datos, suficientemente grande para contener la información significativa, y lo suficientemente pequeña para su poderla manipular rápidamente.

#### **4.7.2.2. Exploración**

Esta etapa explora los datos en la búsqueda de tendencias y anomalías imprevistas, con el fin de obtener la comprensión e ideas para la minería de datos.

#### **4.7.2.3. Modificación**

Esta etapa consiste en la modificación de los datos mediante la creación, selección y transformación de las variables para enfocar el proceso de selección de modelo de minería de datos.

#### **4.7.2.4. Modelado**

Esta etapa consiste en el modelado de los datos, en donde las herramientas de software se encargan de realizar una búsqueda completa de combinaciones de datos que permitirán predecir de forma fiable el resultado deseado.

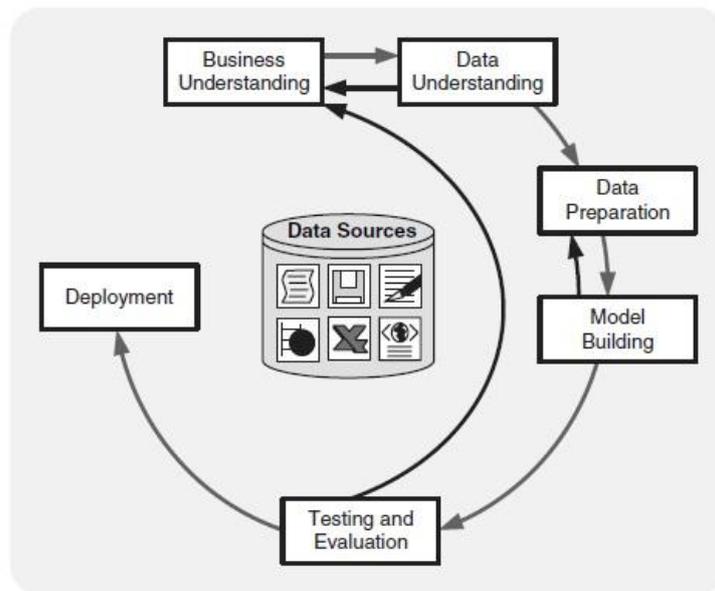
#### **4.7.2.5. Evaluación**

Esta etapa consiste en la evaluación de los datos mediante el análisis de la utilidad y fiabilidad de los resultados del proceso de minería de datos, así como de la estimación de su buen funcionamiento.

Aunque el proceso de SEMMA es independiente de la herramienta de DM elegido, está vinculada a SAS Enterprise Software Miner y pretende orientar al usuario en las implementaciones de aplicaciones de MS. SEMMA ofrece un fácil entender el proceso, lo que permite un desarrollo organizado y adecuado, y mantenimiento de proyectos de DM. Por lo tanto, confiere una estructura para su concepción, creación y evolución, ayudando a presentar soluciones a los problemas de negocios, así como para encontrar de los objetivos de negocio de marcos alemanes (Azevedo & Santos, 2008)

#### **4.7.3. Proceso CRISP-DM**

El proceso CRISP-DM fue desarrollado por medio de los esfuerzos de un consorcio compuesto inicialmente por DaimlerChrysler, SPSS y NCR. CRISP-DM significa (Cross-Industry Standard Process for Data Mining) y consiste en un ciclo que consta de seis etapas (Azevedo & Santos, 2008).



**Figura 12.** Ciclo de Vida del proceso CRISP-DM.

Fuente: (Ncr et al., 2000).

#### **4.7.3.1. Compresión del Negocio.**

Esta fase inicial se centra en la comprensión de los objetivos y requisitos del proyecto desde una perspectiva empresarial, para luego convertir este conocimiento en una definición del problema de minería de datos y un plan preliminar diseñado para alcanzar los objetivos (Azevedo & Santos, 2008).

#### **4.7.3.2. Compresión de los Datos.**

La fase de comprensión de datos comienza con una recolección inicial de los datos, con la finalidad de establecer un primer contacto con el problema, con el fin de familiarizarse con los datos, identificar los problemas de su calidad, y descubrir las primeras ideas sobre los datos, o para detectar subconjuntos interesantes con la finalidad de generar conocimiento sobre información oculta (Azevedo & Santos, 2008).

#### **4.7.3.3. Preparación de los Datos.**

La fase de preparación de datos abarca todas las actividades para la construcción de los datasets (conjuntos de datos) a partir de los datos brutos iniciales, a los cuales se debe aplicar las tareas de limpieza y transformación de tablas, registros y atributos

para con ello lograr calidad en la información resultante que debe quedar lista para ser ingresada en la herramienta de modelado (Azevedo & Santos, 2008).

#### **4.7.3.4. Modelado.**

En esta fase, se seleccionan y se aplica diversas técnicas de modelización y sus parámetros se calibran a los valores óptimos con el apoyo de una herramienta de minería de datos, misma que debe tener las bondades suficientes para permitir aplicar las técnicas necesarias para la construcción del modelo (Azevedo & Santos, 2008).

#### **4.7.3.5. Evaluación.**

En esta etapa el modelo (o modelos) obtenidos son evaluados más a fondo y los pasos ejecutados para construir el modelo son revisados para estar seguro que logra adecuadamente los objetivos del negocio (Azevedo & Santos, 2008).

#### **4.7.3.6. Despliegue.**

Creación del modelo en general, no es el fin del proyecto. Aunque el propósito del modelo es aumentar el conocimiento de los datos, tendrá que ser organizada y presentada de una manera que el cliente puede utilizar el conocimiento adquirido (Azevedo & Santos, 2008).

## **CAPÍTULO V - MARCO METODOLÓGICO**

## 5. Marco Metodológico

### 5.1. Tipo y Nivel

Por la finalidad que persigue esta investigación, la misma reúne las condiciones metodológicas de una investigación pura, pues con esta se pretende aportar beneficios teóricos respecto a la incidencia del uso del internet en el rendimiento de los estudiantes en la Pontificia Universidad Católica del Ecuador. También si vamos en línea a la prolongación de esta investigación en el tiempo, es considerada de tipo transversal o sincrónica, debido a que es dada con datos tomados en un tiempo específico, y no de un estudio realizado por varios años.

Con relación al nivel de la investigación, viene dado respecto al grado de profundidad y alcance que se pretende dar a la misma; este trabajo reúne las características para considerarse como un estudio de nivel correlacional y explicativo, ya que en el mismo se trabajó en la relación entre variables y la explicación de los resultado obtenidos producto de dichas relaciones.

### 5.2. Método y Diseño

Los métodos que se están aplicando en esta investigación son de tipo cuantitativo respecto a los resultados de las relaciones aplicadas entre las variables y los análisis cluster de las mismas.

El diseño que se tomará para este trabajo de investigación se explica por sí solo mediante lo detallado en la [tabla 2]:

**Tabla 2.** Diseño aplicado al trabajo de investigación

INICIO	PREGUNT.	HIPOT.	RESULTADOS	TEORIAS	FIN
Interrogante Inicial	Pregunta 1	Hipótesis 1	Resultado 1	Teoría 1	Conclusión Final
		Hipótesis 2	Resultado 2		
	Pregunta 2	Hipótesis 3	Resultado 3	Teoría 1	
		Hipótesis 4	Resultado 4		

**Fuente:** Realización propia

### **5.3. Población y Muestra.**

La población tomada para esta investigación está conformada por el cuerpo estudiantil de la Pontificia Universidad Católica del Ecuador – Sede Ibarra (PUCE-SI) la cual cuenta con alrededor de 2300 estudiantes matriculados actualmente, de los cuales se levantó encuestas a una muestra significativa de 496 estudiantes que corresponde al 22% de la población total estudiantil.

En la muestra de estudiantes encuestados se tomó estudiantes de todas las carreras que oferta actualmente la Universidad con la finalidad de lograr una interpretación global de los datos obtenidos por cada uno en las encuestas levantadas, dicha muestra está integrada por el 53.80% de hombres y el 46.20% de mujeres.

### **5.4. Técnicas e Instrumentos de recolección de datos.**

En relación a la recolección de datos, fue necesaria la aplicación de encuestas a los estudiantes como ya se lo ha mencionado en el *[Capítulo 5, punto 5.3 Población y Muestra.]*, y para el diseño de este instrumento tanto de fondo como de forma se consideró como referencia al cuestionario empleado en el trabajo de investigación doctoral emprendido por (Torres, 2012). El cuestionario levantado se orientó hacia las siguientes aristas que incide en el estudiante frente al uso del internet:

- Especialidad académica del estudiante
- Información socio-demográfica de los estudiantes mediante tres variables: edad, género e ingresos económicos.
- Información sobre la interacción del estudiante con el internet, a través de las variables: lugar de conexión habitual, días de conexión a la semana, horas de conexión al día, años de conexión y el nivel de conocimiento en su manejo.
- Información sobre la utilización de internet para actividades académicas, cuantificando las mismas respecto al reporte mensual de: Número de veces que se ingresan a la Plataforma virtual, número de consultas al profesor, número de consultas a sus compañeros, número de recursos descargados desde la plataforma virtual, número de videos académicos que ve por medio de youtube, número de foros en los que participa, número de post y tweets

realizados sobre temas académicos, horas de chat sobre temas académicos, horas de búsqueda de información académica en internet, horas de uso de biblioteca virtual de la Universidad.

- Información sobre la utilización de internet para entretenimiento, a través de las siguientes variables: horas a la semana de chat por diversión, horas a la semana de utilización de redes sociales, horas a la semana de utilización de juegos en línea, horas a la semana de descarga desde internet música, videos y programas, número de videos que mira por youtube para entretenimiento semanalmente.
- Medición de interacción de los estudiantes con redes sociales twitter, facebook y LinkedIn. Y administración de contenido en la red mediante blog, youtube y del.icio.us
- Información de acceso al uso de dispositivos electrónicos como: Smartphone, móvil con acceso a internet, Móvil sin acceso a internet, Computador portátil, Tablet, Cámara digital e Ipod.
- Información sobre aprovechamiento del internet por parte de los profesores, tanto en el trabajo de aula como en trabajos enviados a casa.
- Información sobre el rendimiento académico de los estudiantes, mediante datos de número de asignaturas en las que se matriculó cada uno respectivamente en el semestre anterior, y el número de materias aprobadas en el mismo semestre.

## **5.5. Análisis de datos**

Para el análisis de datos se proyectó aplicar técnicas de minería de datos para lograr un modelo descriptivo, que nos ayudaran a confirmar o anular las hipótesis plateadas, mismas que se encuentran detallados en el *[Capítulo 1, punto 1.4 Preguntas de investigación e hipótesis respectivas]* de este documento.

Para lograr un modelo descriptivo robusto y eficaz, se necesitó aplicar algunas técnicas de minería de datos; en el caso de este proyecto han sido escogidas las que

mejor se apegan al objetivo esperado. A continuación detallo las técnicas que han sido seleccionadas para este trabajo de investigación:

### **5.5.1. Análisis Factorial**

El análisis factorial es una técnica estadística que permite la reducción de datos mediante la correlación de un conjunto de variables dadas, obteniendo como resultado un menor número de variables nuevas llamadas factores.

El análisis factorial puede ser de tipo exploratorio o confirmativo. Se considera exploratorio cuando no se conocen los factores correlacionados y por tanto se determinan como parte del resultado del mismo análisis; en cambio el análisis factorial confirmativo se aplica cuando ya se conoce los factores con los que se quiere trabajar respecto a la correlación de las variables, y se confirma dicha correlación de factores a través de los análisis de comprobación o prueba de relación.

En línea a lo comentado anteriormente, esta técnica permitirá obtener factores respecto a variables con las que trabajaremos para lograr el objetivo de la investigación.

### **5.5.2. Análisis Cluster**

También conocido como análisis de conglomerados, es una técnica estadística cuya finalidad es agrupar variables o datos de acuerdo a la similitud que presenten los mismos, logrando entre los grupos confirmados la mayor distinción posible.

El análisis Cluster y el análisis factorial son similares como técnicas de agrupación de variables, sin embargo se diferencian porque el análisis cluster es más flexible y permite trabajar con variables categóricas.

Dentro del análisis cluster se puede optar por dos tipos de análisis: el análisis cluster jerárquico que permite determinar el número óptimo de grupos existentes en los datos; y el análisis cluster K-medias que permite procesar k agrupaciones pero se limita a utilizar un solo método de agrupación, solicitando como información preliminar el número de cluster que se desea obtener. Para este trabajo de investigación se utilizara el tipo k-medias, por las bondades que nos ofrece el mismo.

### **5.5.3. Análisis de regresión Logística**

La regresión logística permite predecir el resultado de una variable categórica, en función de variables predictoras. La probabilidad debería estar entre 0 y 1, y la probabilidad puede ser creciente o decreciente en función a los valores de la variable.

Existen varios tipos de regresión logística, pero los principales son dos:

Modelo logístico univariante o binario. Se utiliza cuando la variable dependiente es una variable binaria, es decir, de solo dos categorías, también conocidas como dummy o dicotómica. Ejemplo: 1 = Sí, 0 = No.

Modelo logístico multivariante o multinomial. Se utiliza cuando la variable dependiente es una variable nominal de más de 2 categorías, de ahí su nombre multinomial. Ejemplo: 1-Soltero, 2-Casado, 3-Divorciado, 4-Viudo 5-Otros.

## **5.6. Definición de variables a utilizar**

Alineados a las hipótesis que mediante este trabajo de investigación se pretende demostrar, se ha tomado las variables con miras a solventar los siguientes ítems:

- Nivel de Ingreso de los estudiantes.
- Usos del internet en actividades académicas.
- Usos del internet en actividades de entretenimiento.
- Rendimiento académico de los estudiantes.

Y las variables seleccionadas preliminarmente por estar relacionadas a los ítems presentados en líneas anteriores son:

### **5.6.1. Variable de nivel de ingreso**

Para medir el nivel de ingreso de los estudiantes se tomara la variable recabada a través de la pregunta cinco de la encuesta levantadas a los estudiantes de nivel superior, mismo que presenta cinco segmentos ya definidos desde la encuesta como tal.

### **5.6.2. Variables de Usos de Internet en actividades académicas.**

Las variables a considerar para lograr medir los usos del internet en actividades académicas son las diez consultas levantadas en las encuestas a los estudiantes, y están agrupadas en la pregunta once de la encuesta [Ver Anexo 1].

El procedimiento para obtener las variables elaboradas respecto al uso de internet en actividades académicas se divide en tres fases: reducción de variables a través del análisis factorial; luego la clasificación y verificación utilizando los factores obtenidos en el análisis factorial como información de inicial para aplicar el análisis cluster; y finalmente la interpretación de grupos, en los cuales se muestra los cluster resultantes.

### **5.6.3. Variables de Usos de Internet en actividades de entretenimiento.**

Las variables a considerar para lograr medir los usos del internet en actividades de entretenimiento son las cinco consultas levantadas en las encuestas a los estudiantes, están agrupadas en la pregunta doce de la encuesta [Ver Anexo 1].

El procedimiento para lograr obtener las variables elaboradas respecto al uso de internet en actividades de entretenimiento se divide en tres fases: reducción de variables a través del análisis factorial; luego la clasificación y verificación utilizando los factores obtenidos en el análisis factorial como información de inicial para aplicar el análisis cluster; y finalmente la interpretación de grupos, en los cuales se muestra los cluster resultantes.

### **5.6.4. Variables de rendimiento académico**

Las variables a considerar para lograr obtener el rendimiento académico de los estudiantes son las dos consultas levantadas en las encuestas, agrupadas en la pregunta dieciocho de la encuesta aplicada [Ver Anexo 1].

Para lograr obtener las variables elaboradas respecto al rendimiento académico de los estudiantes se aplicara la técnica de “Regla de tres” en donde **A** será el valor levantado en la encuesta respecto al número de asignaturas en las que se matriculo relacionado con **B** que expresa el porcentaje que en este caso sería 100%, y **X** será el valor levantado en la encuesta respecto al número de asignaturas aprobadas relacionado con **Y** que sería el valor resultante. La expresión de la ecuación sería:

$$\left. \begin{array}{l} A \longrightarrow B \\ X \longrightarrow Y \end{array} \right\} \rightarrow Y = \frac{B \cdot X}{A}$$

### 5.7. Herramienta de minería de datos a utilizar

Luego de haber analizado las bondades que nos ofrecen las principales herramientas de minería de datos, se ha seleccionado para el desarrollo de este trabajo de investigación dos de estas herramientas.

La primera es SPSS, que será utilizada para el ejercicio de exploración de datos, misma que nos permitirá conocer estadísticas esenciales de los datos levantados por medio de las encuestas y poder definir del universo de variables, los DataSet que se necesitan para poder crear el modelo de minería de datos. El motivo por el que se escogió a esta herramienta es su facilidad y flexibilidad para mostrar modelos estadísticos, así como la ventaja adicional de que permite visualizar histogramas y diagramas de dispersión, necesarios para la toma de decisiones respecto a cada variable.

La segunda herramienta seleccionada se utilizará para la implementación de las técnicas de minería de datos en este trabajo de investigación, y será WEKA (*Waikato Environment for Knowledge Analysis*).

Como ya se había mencionado en el [CAPITULO 04, punto 4.6], Weka (*Waikato Environment for Knowledge Analysis*) es una herramienta tecnológica multiplataforma de libre distribución bajo licencia GNU-GPL que fue desarrollada por los investigadores de la Universidad de Waikato en Nueva Zelanda; permite aplicar técnicas de minería de datos como clustering, regresión, clasificación, entre otros; permite el acceso a datos a través de un archivo en formato ARFF(archivo plano) y posee una interfaz gráfica para el usuario.

#### ¿Por qué utilizar dos herramientas y no solo una?

La razón principal de utilizar dos herramientas, es la de aprovechar las bondades predominantes de cada una de estas, en relación a las necesidades que surgen en el proceso de Minería de datos CRISP-DM (Cross-Industry Standard Process for Data Mining), mismo que va a ser aplicado para este trabajo de investigación.

Las ventajas y limitantes más significativas de SPSS son:

#### Ventajas

- ✓ Potente en cálculos estadísticos
- ✓ Permite generar gráficos y tablas estadísticas de una forma muy fácil y personalizada.
- ✓ Utiliza cuadros de dialogo que permiten determinar las acciones y seleccionar los análisis que se necesiten.

#### Desventajas

- ✓ Es limitado en algoritmos de Minería de datos, frente a otras herramientas
- ✓ Es una herramienta pagada.
- ✓ El usuario debe tener experiencia previa utilizando SPSS para que pueda discernir las opciones a utilizar.

Las ventajas y limitantes más significativas de WEKA son:

#### Ventajas

- ✓ Al estar completamente programada en Java permite correr la misma en la mayoría de las plataformas.
- ✓ Contiene muchos algoritmos de técnicas de minería de datos para poder aplicar los diferentes modelos de MD.
- ✓ Proporciona información adicional que permite evaluar el nivel de efectividad de los modelos resultantes.

#### Desventajas

- ✓ Existe poca documentación para el usuario respecto a la utilización de la herramienta.
- ✓ No permite generar gráficos y tablas estadísticas de manera fácil y personalizada.

Con lo antes mencionado la definición de utilizar las dos herramientas se explica por sí sola, pues mientras SPSS nos ayuda fácilmente con temas estadísticos, que en los primeros pasos del proceso CRISP-DM nos resulta muy útil; la herramienta WEKA

lleva la delantera en la aplicación de técnicas y algoritmos de minería de datos, lo cual es de vital importancia en los pasos de desarrollo e implementación del modelo de minería de datos dentro del proceso. En la [Tabla 3] se presenta el cuadro comparativo de los algoritmos que ofrecen algunas herramientas de minería de datos, en donde se ve claramente que la herramienta WEKA tiene la mayor participación que SPSS, lo cual también es un factor importante para la decisión de utilizar WEKA en la implementación del modelo de Minería de Datos.

**Tabla 3.** Cuadro comparativo de algoritmos en herramientas de Minería de Datos

	WEKA	R > 2.7	Tanagra	KNIME	SPSS	RapidMiner
Algoritmos implementados de forma nativa	168	24	13	9	10	34
Algoritmos Importados desde Weka	0	0	0	102	0	101
Total Algoritmos Implementados	168	24	13	111	10	135
% algoritmos nativos	100%	100%	100%	8.11%	100%	25.19%
% algoritmos nativos sobre el total	65.12%	9.30%	5.04%	3.49%	3.88%	13.18%

**Fuente:** Elaboración propia

## **CAPÍTULO VI - DESARROLLO**

## **6. Desarrollo**

Para el desarrollo de este trabajo de investigación se aplicará la metodología de minería de datos CRISP-DM (Cross-Industry Standard Process for Data Mining), misma que ya fue mostrada en el apartado del Marco Teórico [*Capítulo 04 sección 4.7 punto 4.7.3*].

### **6.1. Fase I - Compresión del Negocio.**

#### **6.1.1. Determinación de objetivos del negocio**

La Educación superior en el Ecuador y específicamente en la Pontificia Universidad Católica del Ecuador – Sede Ibarra, busca cultivar en sus estudiantes siempre las actitudes y aptitudes suficientes para que puedan lograr el éxito académico. Así, la PUCE-SI se ha propuesto los siguientes objetivos institucionales en su plan estratégico 2009-2016 (PUCE-SI, 2008):

- Realizar una eficiente gestión administrativa que brinde servicios de calidad.
- Generar una docencia de calidad mediante procesos de enseñanza aprendizaje integral, de orden teórico práctico, acordes a los requerimientos de la sociedad.
- Ofertar nuevas carreras en diferentes modalidades de estudio para satisfacer la demanda social de la región
- Desarrollar una investigación científica y tecnológica pertinente para generar teoría y contribuir a la solución de los problemas socio-económicos.
- Formar investigadores con solvencia ética y científica a nivel docente y discente para planificar, desarrollar y proyectar las investigaciones.

Con estos objetivos está claro que la visión de la PUCE-SI es formar a estudiantes competitivos y potenciales investigadores, a través de una docencia competitiva y orientada a la investigación científica y tecnológica.

Ahora bien, con la inmersión de la tecnología en la sociedad universitaria, con la rápida evolución con la que se está desarrollando la misma, y con la finalidad de hacer prevalecer los objetivos institucionales, se ha visto la necesidad de emprender el presente trabajo de investigación, en donde se busca entender y confirmar la

incidencia que tiene el nivel de ingresos de los estudiantes de la PUCE-SI en el uso del internet, y la repercusión que este uso de internet puede tener en el éxito académico de los estudiantes. Esto toda vez que se ha podido observar estudios similares aplicados en otras instituciones educativas que han presentado, en algunos resultados positivos y en otros resultados negativos.

### **6.1.2. Valoración de la situación**

Con la necesidad o problemática citada en el punto anterior, se ha procedido a valorar la situación actual como input para iniciar el desarrollo del proyecto, permitiéndome a continuación detallar los recursos, requisitos, supuestos, restricciones, costes y beneficios para la ejecución de este proyecto.

#### **6.1.2.1. Recursos**

Recursos – Personas

- ✓ Ing. Franco Guamán – Director del Trabajo de Investigación.
- ✓ PhD. Juan Carlos Torres – Docente Investigador Tecnológico.
- ✓ PhD. Héctor Gómez – Experto en Minería de Datos e IA.
- ✓ Jaime Torres Carrión – Estudiante Investigador Tecnológico.

Recursos Tecnológicos (Software y Hardware)

- ✓ Hardware – Computador Personal o portátil, con procesador Intel Core I5, memoria RAM de 4GB y un Disco Duro de 500GB.
- ✓ Software – Weka 3.6, para la implementación de las técnicas de minerías de datos necesarias para este proyecto.
- ✓ Software – SPSS 14.0 for Windows, para la aplicación de estadísticas básicas la exploración y limpieza de los datos.
- ✓ Software – Bloc de Notas, para crear los DataSet necesarios para la herramienta Weka en formato \*.arff
- ✓ Software – Microsoft Excel, para tabulación de la información levantada en las encuestas.
- ✓ Software – Microsoft Word, para documentación del proceso de análisis, resultados obtenidos e indicadores.

- ✓ Software – Mendeley, para la búsqueda de información científica y administración de bibliografía.

#### Recursos – Datos

- ✓ Datos socio-económica de los estudiantes.
- ✓ Datos referentes al uso del internet para actividades académicas.
- ✓ Datos referentes al uso del internet para actividades de entretenimiento.
- ✓ Datos de rendimiento académico (asignaturas matriculadas y asignaturas aprobadas en el último semestre terminado).

#### **6.1.2.2. Requerimientos.**

Se ha identificado los siguientes requerimientos para el desarrollo de este proyecto.

- ✓ Contar con información de los estudiantes suficiente para realizar la aplicación de técnicas de minería de datos y generar un modelo de calidad.
- ✓ Contar con la participación activa de las personas detalladas en el punto anterior, con la finalidad discutir resultados y concordar en las conclusiones.
- ✓ Dominar las herramientas tecnológicas necesarias para la ejecución de este proyecto. Mismas que ya fueron detalladas en el punto anterior.

#### **6.1.2.3. Supuestos.**

- ✓ Si un estudiante se matricula en **n** materias, máximo aprobará **n** materias.
- ✓ Un estudiantes universitario no debería tener una edad menor a 16 años, esto por lo años de estudio primario y secundario que tiene que haber aprobado antes de llegar a la Universidad.
- ✓ Se supone que se cuenta con información real y confiable.

#### **6.1.2.4. Restricciones**

Para la generación del modelo de minería de datos se está tomando información proporcionada por lo estudiantes mediante encuestas aplicadas, en el caso de que la información este incompleta se procederá con el proceso de limpieza de datos.

#### **6.1.2.5. Costes y beneficios**

Ya que el presente proyecto de investigación se está realizando como trabajo de fin de carrera de estudios de pregrado de la UTPL, los costes del mismo serán asumidos por el estudiante investigador y no presentara ningún costo para la PUCE-SI.

El beneficio principal de este proyecto es la confirmación o negación de las hipótesis planteadas, basándose para ello en los resultados e indicadores sustentados en esta investigación.

### 6.1.3. Determinación de objetivos de Minería de Datos

- ✓ Relacionar variables de nivel de ingresos, usos de internet (para actividades académicas y actividades de entretenimiento) y el rendimiento o éxito académico de los estudiantes.
- ✓ Identificar grupos de estudiantes, según atributos comunes relacionados al uso de internet en actividades académicas.
- ✓ Identificar grupos de estudiantes, según atributos comunes relacionados al uso de internet en actividades de entretenimiento.
- ✓ Calcular variables de rendimiento académico a partir de los datos de materias inscritas y materias aprobadas por cada estudiante.

### 6.1.4. Creación del plan del proyecto

El plan de ejecución ha sido realizado pensando en la secuencia de fases y temas que cada una de ellas requiere para asegurar la calidad del modelo de minería de datos [Ver tabla 4].

**Tabla 4.** Plan de Ejecución del Proyecto

FASE CRISP-DM	TEMAS A REALIZAR	TIEMPO ESTIMADO (EN SEMANAS)
Comprensión del Negocio	Objetivos del Negocio y Valoración de la situación	1
	Recursos, requerimientos, supuestos, restricciones, costes y beneficios.	1
Comprensión de los datos	Recolección de datos iniciales (incluye levantamiento de encuestas y tabulación de las mismas).	6
	Descripción de los datos	1

	Exploración de datos.	2
	Verificación de la calidad de los datos	1
Preparación de los datos	Selección de datos	1
	Limpieza de los datos	1
	Estructuración de los datos	1
	Integración de los datos	1
	Formateo de los datos	1
Modelado	Seleccionar técnica de modelado	1
	Construir el modelo	4
Evaluación	Evaluación de los resultados	1
	Revisión del proceso	1
Despliegue o Implementación	Informe Final	1
	Revisión del proyecto	1

Fuente: Elaboración propia

## 6.2. Fase II - Compresión de los datos

### 6.2.1. Recolección de los datos iniciales

Los datos iniciales o de origen de la información para este trabajo de investigación se lograron mediante la aplicación de encuestas físicas a una muestra de 500 estudiantes de la PUCE-SI de diferentes niveles de educación y de diferentes carreras que oferta la universidad. El formato aplicado en cada encuesta se puede visualizar en el [Anexo 1].

Al haberse realizado las encuestas de forma impresa, se tuvo que aplicar un proceso de digitalización de la información, tomándose las debidas precauciones para conservar la integridad de la información.

### 6.2.2. Descripción de los datos.

## Descripción de datos iniciales

Mediante el cuestionario aplicado a los estudiantes se logró recabar la información inicial para el trabajo de investigación, la cual se encuentra tabulada en un archivo en formato Excel con cada uno de los campos debidamente etiquetados [Ver Tabla 5]:

**Tabla 5.** Información captada mediante encuestas a estudiantes Universitarios

COD_CAMPO	PREGUNTA	DETALLE DEL CAMPO
uni	¿En qué universidad estudia?	En este caso al haber aplicado este trabajo a una sola Universidad, todas tendrán como dato PUCE-SI
car	¿Qué carrera estudia?	Pueden ser: Jurisprudencia, Informática, Arquitectura, Comercio Exterior, Agropecuaria, Diseño, entre otras
eda	¿Cuál es su edad?	Este dato siempre va a ser numérico y debe tener cifras entre 16 y 65 años.
sex	¿Cuál es su género?	En este campo se registrara H = Hombre y M = Mujer.
ing	Los ingresos mensuales de su familia	Este campo registrara uno de los siguientes valores: 1 = Hasta 350 dólares 2 = Hasta 600 dólares 3 = Hasta 1.000 dólares 4 = Hasta 1.500 dólares 5 = Más de 1.500 dólares
lug_con	¿Desde dónde se conecta habitualmente a Internet? (escoja solo una opción)	Podrá registrar uno de los siguientes valores : 1 = Desde la casa. 2 = Desde el cyber café. 3 = Desde el trabajo. 4 = Desde la Universidad. 5 = Red móvil (movistar, claro, cnt).
dia_con	De 1 a 7, ¿cuántos días a la semana se conecta Internet?	Este campo podrá tener registrado un valor numérico contenido entre el 1 y el 7.
niv_con	De 1 a 10 su nivel de conocimientos en el manejo de Internet es:	Podrá registrar un valor numérico contenido entre el 1 y el 10 (siendo 1 el conocimiento más bajo y 10 el conocimiento más avanzado).
hor_dia	¿Aproximadamente cuántas horas se conecta cada día?	Este campo podrá tener registrado un valor numérico contenido entre el 0 y el 24.
año_ext	¿Hace cuántos años se conecta a Internet?	Campo que podrá tener registrado un valor numérico no mayor al de la edad del estudiante.
a_ing_pla	¿Cuántas veces a la semana ingresa a la plataforma virtual de su universidad?	Este campo siempre estará poblado por un valor numérico.
a_con_pro	¿Aproximadamente cuántas consultas les hace a sus profesores cada mes?	Campo que siempre será un valor numérico.

a_con_est	¿Aproximadamente cuántas consultas realiza a sus compañeros por mes?	Campo que siempre será un valor numérico.
a_rec_edu	¿Aproximadamente cuántos recursos educativos descarga de la plataforma virtual cada mes?	Campo que siempre será un valor numérico.
a_vid_aca	¿Aproximadamente cuántos videos académicos mira en youtube cada mes?	Campo que siempre será un valor numérico.
a_for_vir	¿Aproximadamente en cuántos foros virtuales participa cada mes?	Campo que siempre será un valor numérico.
a_pos_aca	¿Aproximadamente cuántos post o tweets sobre temas académicos realiza en las redes sociales por mes?	Campo que siempre será un valor numérico.
a_hor_cha	¿Aproximadamente cuántas horas chatea sobre temas académicos cada mes?	Campo que siempre será un valor numérico.
a_bus_inf	¿Aproximadamente cuántas horas busca información académica en internet cada mes?	Campo que siempre será un valor numérico.
a_bib_vir	¿Aproximadamente cuántas horas utiliza la biblioteca virtual de la universidad cada mes?	Campo que siempre será un valor numérico.
e_hor_cha	¿Aproximadamente cuántas horas a la semana chatea por diversión?	Campo que siempre será un valor numérico.
e_red_soc	¿Aproximadamente cuántas horas a la semana utiliza redes sociales?	Campo que siempre será un valor numérico.
e_jue_lin	¿Aproximadamente cuántas horas a la semana utiliza juegos en línea?	Campo que siempre será un valor numérico.
e_des_con	¿Aproximadamente cuántas horas a la semana descarga música, videos y programas?	Campo que siempre será un valor numérico.
e_vid_sem	¿Aproximadamente cuántos videos para entretenimiento mira en youtube cada semana?	Campo que siempre será un valor numérico.
rs_seg_twi	¿Aproximadamente cuántos seguidores tiene	Campo que siempre será un valor numérico.

	en twitter?	
rs_seg_fac	¿Cuántos amigos tiene en facebook?	Campo que siempre será un valor numérico.
rs_seg_lin	¿Cuántos contactos tiene en LinkedIn?	Campo que siempre será un valor numérico.
cc_blog	Tiene un blog	Este campo podrá registrar 1 = Si y 2 = No.
cc_youtub	Tiene cuenta en youtube	Este campo podrá registrar 1 = Si y 2 = No.
cc_delicius	Tiene cuenta en www.del.icio.us	Este campo podrá registrar 1 = Si y 2 = No.
d_tel_cam	Smartphone con cámara fotográfica y acceso a internet	Este campo podrá tener registrado un valor numérico contenido entre el 1 y el 10 (siendo 1 el uso más bajo y 10 el uso más avanzado).
d_tel_int	Teléfono móvil con acceso a internet	Este campo podrá tener registrado un valor numérico contenido entre el 1 y el 10 (siendo 1 el uso más bajo y 10 el uso más avanzado).
d_tel	Teléfono móvil sin acceso a internet	Este campo podrá tener registrado un valor numérico contenido entre el 1 y el 10 (siendo 1 el uso más bajo y 10 el uso más avanzado).
d_com_por	Computador portátil	Dato de nivel de uso del computador portátil, este campo podrá tener registrado un valor numérico contenido entre el 1 y el 10 (siendo 1 el uso más bajo y 10 el uso más avanzado).
d_tab	Tablet (iPad; Galaxy Tab, Kindle, etc)	Dato de nivel de uso de Tablet, este campo podrá tener registrado un valor numérico contenido entre el 1 y el 10 (siendo 1 el uso más bajo y 10 el uso más avanzado).
d_cam_dig	Cámara digital	Dato de nivel de uso de Cámara digital, este campo podrá tener registrado un valor numérico contenido entre el 1 y el 10 (siendo 1 el uso más bajo y 10 el uso más avanzado).
d_ipod	iPod / MP3 Player	Dato de nivel de uso de Ipod / MP3 player, este campo podrá tener registrado un valor numérico contenido entre el 1 y el 10 (siendo 1 el uso más bajo y 10 el uso más avanzado).
pe_rap_esf	Internet le permite elaborar los trabajos más rápido y con menos esfuerzo	Dato de nivel de ayuda que ofrece el internet para elaborar trabajos con mayor facilidad, este campo podrá tener registrado un valor numérico contenido entre el 1 y el 10 (siendo 1 el uso más bajo y 10 el uso más avanzado).
pe_conf_inf	Usted confía en la información de internet para realizar sus tareas	Dato de nivel de confianza que le ofrece el internet al estudiante de certeza de información, este campo podrá tener registrado un valor numérico contenido entre el 1 y el 10 (siendo 1 el uso más bajo y 10 el uso más avanzado).
pe_pro_apr	Internet le permite prescindir de la Biblioteca	Dato de nivel de certeza que el internet le da al estudiante para prescindir de la biblioteca, este campo podrá tener registrado un valor numérico contenido entre el 1 y el 10 (siendo 1 el uso más bajo y 10 el uso más avanzado).

pe_fac_apr	Internet facilita el proceso de aprendizaje	Nivel de certeza que el internet brinda para facilitar al estudiante el proceso de aprendizaje, este campo podrá registrar un valor numérico contenido entre el 1 y 10 (siendo 1 el uso más bajo y 10 el uso más avanzado).
pe_mej_cal	Internet le permite mejorar sus calificaciones	Nivel de certeza que el internet da al estudiante para mejorar sus calificaciones, este campo podrá tener registrado un valor numérico contenido entre el 1 y el 10 (siendo 1 el uso más bajo y 10 el uso más avanzado).
pe_tra_int	Usted presenta trabajos académicos copiados desde Internet	Porcentaje en los que el estudiante considera copia trabajos académicos, este campo podrá tener registrado un valor numérico contenido entre el 1 y el 10 (siendo 1 el uso más bajo y 10 el uso más avanzado).
pr_pla_vir	Su profesor ingresa a la plataforma virtual	Dato de si su profesor ingresa a la plataforma virtual, se podrá registrar 1 = Si, 2 = No y 3 = A veces.
pr_con_cor	Contesta sus consultas por correo electrónico	Este campo podrá registrar 1 = Si, 2 = No y 3 = A veces.
pr_cha_aca	Chatea con usted eventualmente sobre aspectos académicos	Dato de si su profesor chatea con usted eventualmente sobre aspectos académicos, este campo podrá registrar 1 = Si, 2 = No y 3 = A veces.
pr_red_tem_aca	Su profesor comenta en redes sociales sobre temas académicos	Este campo podrá registrar 1 = Si, 2 = No y 3 = A veces.
pr_mat_dig_pla	Le sube materiales digitales a la plataforma virtual	Este campo podrá registrar 1 = Si, 2 = No y 3 = A veces.
pr_rec_bib_vir	Le recomienda recursos digitales de la biblioteca virtual	Este campo podrá registrar 1 = Si, 2 = No y 3 = A veces.
pr_vid_aca	Le recomienda videos sobre temas académicos	Este campo podrá registrar 1 = Si, 2 = No y 3 = A veces.
pr_eva_pla_vir	Le plantea cuestionarios o evaluaciones en la plataforma virtual	Este campo podrá registrar 1 = Si, 2 = No y 3 = A veces.
pr_for_vir	Le plantea foros virtuales	Este campo podrá registrar 1 = Si, 2 = No y 3 = A veces.
pr_tie_pag_web	Su profesor tiene una página web, blog o perfil de facebook	Este campo podrá registrar 1 = Si, 2 = No.
pr_tie_cue_twi	Su profesor tiene cuenta de twitter	Este campo podrá registrar 1 = Si, 2 = No.
as_mat_sem	En el semestre anterior, ¿en cuántas asignaturas se matriculó?	Este campo procede a registrar siempre valores numéricos.
as_apr_sem	En el semestre anterior ¿cuántas asignaturas aprobó?	Este campo procede a registrar siempre valores numéricos.

---

**Fuente:** Elaboración propia

### Descripción de DataSet seleccionados

Para el presente trabajo de investigación se han seleccionado cuatro DataSet, los cuales abarcan las cuatro hipótesis de estudio.

A continuación tenemos un detalle de los atributos que conforman cada DataSet con el fin de ser utilizados para la generación del modelo.

✓ *DataSet – Ingreso Económico del estudiante*

Contiene los datos de ingreso familiar del estudiante, este DataSet está compuesto de un campo **ing**, mismo que podrá tener registrado uno de los siguientes valores de acuerdo al rango ingresado por el estudiante en la encuesta:

- 1 = Hasta 350 dólares
- 2 = Hasta 600 dólares
- 3 = Hasta 1.000 dólares
- 4 = Hasta 1.500 dólares
- 5 = Más de 1.500 dólares

Este DataSet será relacionado con los DataSet *Uso internet en actividades de académicas* y *Uso internet en actividades de entretenimiento*, para la generación del modelo.

✓ *DataSet – Uso internet en actividades académicas*

Contiene los datos de las diferentes formas de uso del internet en actividades académicas por parte del estudiante, este DataSet está compuesto de los campos detallados en [Tabla 6].

**Tabla 6.** Uso internet en actividades académicos - DataSet

COD_CAMPO	DETALLE DE PREGUNTA GUARDADA EN CADA CAMPO DEL DATASET
a_ing_pla	¿Cuántas veces a la semana ingresa a la plataforma virtual de su universidad?
a_con_pro	¿Aproximadamente cuántas consultas les hace a sus profesores cada mes?
a_con_est	¿Aproximadamente cuántas consultas realiza a sus compañeros por mes?
a_rec_edu	¿Aproximadamente cuántos recursos educativos descarga de la plataforma virtual cada mes?
a_vid_aca	¿Aproximadamente cuántos videos académicos mira en youtube cada mes?
a_for_vir	¿Aproximadamente en cuántos foros virtuales participa cada mes?
a_pos_aca	¿Aproximadamente cuántos post o tweets sobre temas académicos realiza en las redes sociales por mes?
a_hor_cha	¿Aproximadamente cuántas horas chatea sobre temas académicos cada mes?
a_bus_inf	¿Aproximadamente cuántas horas busca información académica en internet cada mes?
a_bib_vir	¿Aproximadamente cuántas horas utiliza la biblioteca virtual de la universidad cada mes?

**Fuente:** Elaboración propia

Este DataSet será relacionado con los DataSet *Ingreso económico del estudiante* y *Rendimiento académico*, para la generación del modelo.

✓ *DataSet – Uso internet en actividades de entretenimiento.*

Contiene los datos de las diferentes formas de uso del internet en actividades de entretenimiento por parte del estudiante, este DataSet está compuesto de los campos detallados en la [tabla 7].

**Tabla 7.** Uso internet en actividades de entretenimiento – DataSet

COD_CAMPO	DETALLE DE PREGUNTA GUARDADA EN CADA CAMPO DEL DATASET
e_hor_cha	¿Aproximadamente cuántas horas a la semana chatea por diversión?
e_red_soc	¿Aproximadamente cuántas horas a la semana utiliza redes sociales?
e_jue_lin	¿Aproximadamente cuántas horas a la semana utiliza juegos en línea?
e_des_con	¿Aproximadamente cuántas horas a la semana descarga música, videos y programas?
e_vid_sem	¿Aproximadamente cuántos videos para entretenimiento mira en youtube cada semana?

**Fuente:** Elaboración propia

Este DataSet será relacionado con dos DataSet *Ingreso económico del estudiante* y *Rendimiento académico*, para la generación del modelo.

✓ *Rendimiento académico*

Contiene los datos de número de materias en las que se matriculo el estudiante encuestado, y el número de materias que aprobó, este DataSet está compuesto de los campos detallados en [Tabla 8].

**Tabla 8.** Rendimiento académico – DataSet

COD_CAMPO	DETALLE DE PREGUNTA GUARDADA EN CADA CAMPO DEL DATASET
as_mat_sem	En el semestre anterior, ¿en cuántas asignaturas se matriculó?
as_apr_sem	En el semestre anterior ¿cuántas asignaturas aprobó?

**Fuente:** Elaboración propia

Este DataSet será relacionado con los DataSet *Uso internet en actividades de académicas* y *Uso internet en actividades de entretenimiento*, para la generación del modelo.

### 6.2.3. Exploración de datos.

Luego de haber descrito los datos iniciales, se procede a explorar los mismos, esto respecto a los campos que se tomaron para formar los DataSet que nos ayudaran a la generación del modelo, estos son *Ingresos económicos*, *Uso Internet para actividades académicas*, *Uso internet para actividades de entretenimiento* y *Rendimiento académico*. Para este ejercicio de exploración de datos aplicando estadística se utilizará la herramienta SPSS, esto en línea a lo ya explicado en [Capítulo 5, Sección 5.7], y los entregables que se presentara serán la tabla de frecuencia, gráfico de barras de distribución y cuadro estadístico con las operaciones moda, media y mediana; esto nos permitirá tener un primer panorama del estado de cada variable a utilizar respecto a los datos.

## 1. DataSet - Ingresos económicos de los estudiantes

**Tabla 9.** Frecuencias de ingresos económicos

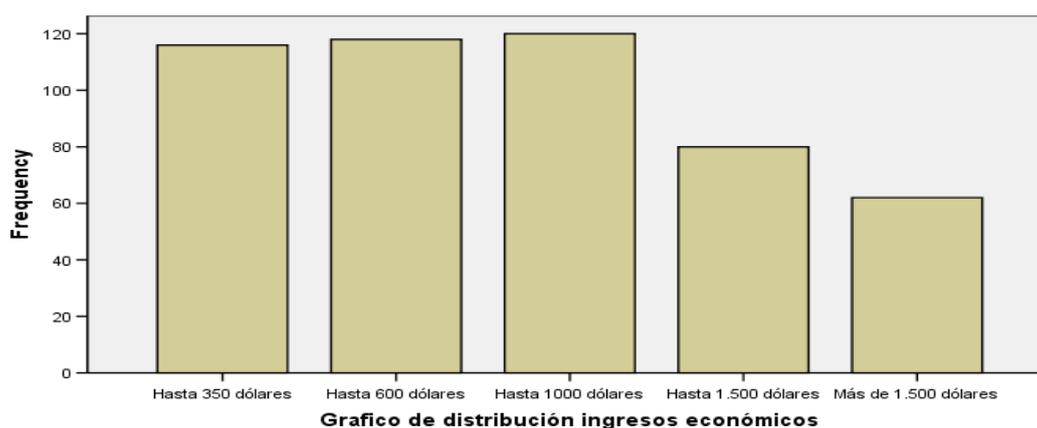
	Frecuencia	Porcentaje	Porcentaje Valido	Porcentaje Acumulativo
Hasta 350 dólares	116	23.40	23.40	23.40
Hasta 600 dólares	118	23.80	23.80	47.20
Hasta 1000 dólares	120	24.20	24.20	71.40
Hasta 1.500 dólares	80	16.10	16.10	87.50
Más de 1.500 dólares	62	12.50	12.50	100.00
Total	496	100.00	100.00	

**Fuente:** Elaboración propia a través de Herramienta SPSS

**Tabla 10.** Operaciones Estadísticas de ingresos económicos

Número de registros	Validos	496
	Desaparecidos	0
Media		2.71
Mediana		3.00
Moda		3.00

**Fuente:** Elaboración propia a través de Herramienta SPSS



**Figura 13.** Grafica de distribución Ingresos Económicos

**Fuente:** Generado desde herramienta SPSS

## 2. DataSet - Usos internet para actividades académicas.

¿Cuántas veces a la semana ingresa a la plataforma virtual de su universidad?

**Tabla 11.** Frecuencias de ¿Cuántas veces a la semana ingresa a la plataforma virtual de su universidad?

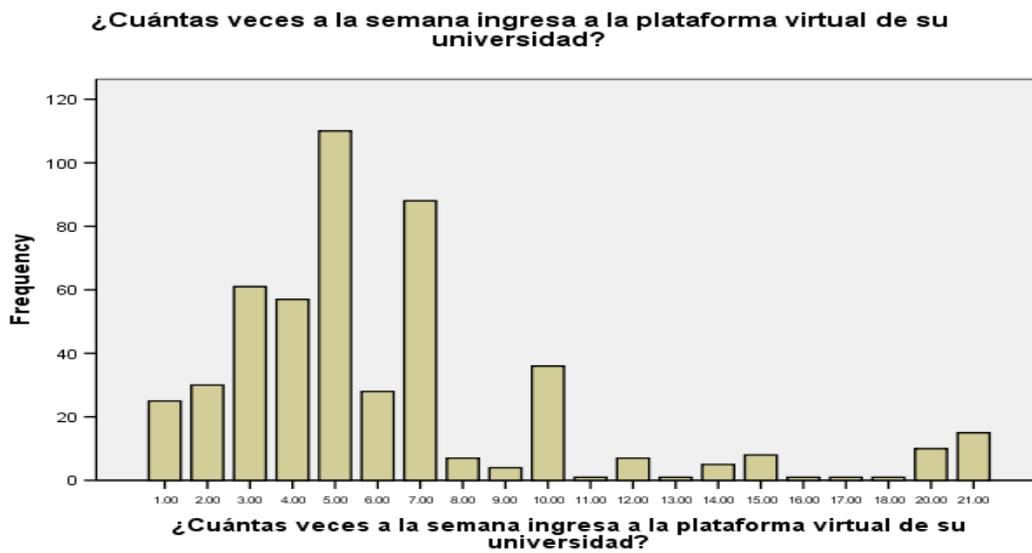
	Frecuencia	Porcentaje	Porcentaje Valido	Porcentaje Acumulativo
1	25	5.00	5.00	5.00
2	30	6.00	6.00	11.10
3	61	12.30	12.30	23.40
4	57	11.50	11.50	34.90
5	110	22.20	22.20	57.10
6	28	5.60	5.60	62.70
7	88	17.70	17.70	80.40
8	7	1.40	1.40	81.90
9	4	0.80	0.80	82.70
10	36	7.30	7.30	89.90
11	1	0.20	0.20	90.10
12	7	1.40	1.40	91.50
13	1	0.20	0.20	91.70
14	5	1.00	1.00	92.70
15	8	1.60	1.60	94.40
16	1	0.20	0.20	94.60
17	1	0.20	0.20	94.80
18	1	0.20	0.20	95.00
20	10	2.00	2.00	97.00
21	15	3.00	3.00	100.00
<b>Total</b>	496	100.00	100.00	

**Fuente:** Elaboración propia a través de Herramienta SPSS

**Tabla 12.** Operaciones Estadísticas ¿Cuántas veces a la semana ingresa a la plataforma virtual de su universidad?

Número de registros	Validos	496
	Desaparecidos	0
Media		6.34
Mediana		5.00
Moda		5.00

**Fuente:** Elaboración propia a través de Herramienta SPSS



**Figura 14.** Grafica de distribución de ¿Cuántas veces a la semana ingresa a la plataforma virtual de su universidad?

**Fuente:** Generado desde herramienta SPSS

¿Aproximadamente cuántas consultas a través de TIC's les hace a sus profesores cada mes?

**Tabla 13.** Frecuencias de ¿Cuántas consultas a través de TIC's les hace a sus profesores cada mes?

	Frecuencia	Porcentaje	Porcentaje Valido	Porcentaje Acumulativo
1	45	9.10	9.10	9.10
2	53	10.70	10.70	19.80
3	37	7.50	7.50	27.20
4	40	8.10	8.10	35.30
5	105	21.20	21.20	56.50
6	17	3.40	3.40	59.90
7	13	2.60	2.60	62.50
8	17	3.40	3.40	65.90
9	2	0.40	0.40	66.30
10	80	16.10	16.10	82.50
11	3	0.60	0.60	83.10
12	4	0.80	0.80	83.90
13	4	0.80	0.80	84.70
15	18	3.60	3.60	88.30
16	1	0.20	0.20	88.50
18	4	0.80	0.80	89.30
20	31	6.30	6.30	95.60
22	1	0.20	0.20	95.80
25	7	1.40	1.40	97.20
28	1	0.20	0.20	97.40
30	13	2.60	2.60	100.00
<b>Total</b>	496	100.00	100.00	

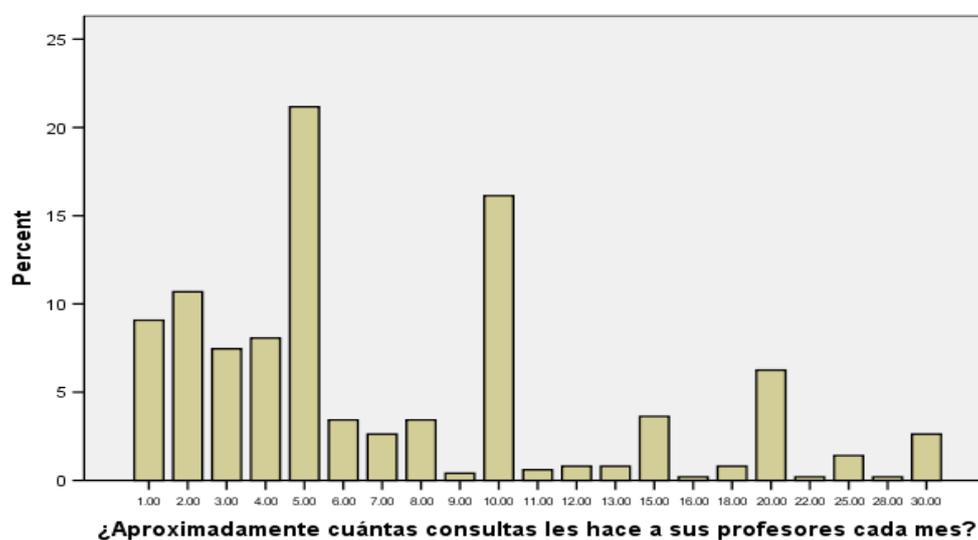
**Fuente:** Elaboración propia a través de Herramienta SPSS

**Tabla 14.** Operaciones Estadísticas ¿Cuántas consultas a través de TIC's les hace a sus profesores cada mes?

<b>Número de registros</b>	Validos	496
	Desaparecidos	0
<b>Media</b>		7.70
<b>Mediana</b>		5.00
<b>Moda</b>		5.00

**Fuente:** Elaboración propia a través de Herramienta SPSS

**¿Aproximadamente cuántas consultas les hace a sus profesores cada mes?**



**Figura 15.** Grafica de distribución de ¿Cuántas consultas a través de TIC's les hace a sus profesores cada mes?

**Fuente:** Generado desde herramienta SPSS

¿Aproximadamente cuántas consultas a través de TIC's realiza a sus compañeros por mes?

**Tabla 15.** Frecuencias de ¿Cuántas consultas a través de TIC's realiza a sus compañeros por mes?

	Frecuencia	Porcentaje	Porcentaje Valido	Porcentaje Acumulativo
1	24	4.80	4.80	4.80
2	30	6.00	6.00	10.90
3	29	5.80	5.80	16.70
4	24	4.80	4.80	21.60
5	81	16.30	16.30	37.90
6	16	3.20	3.20	41.10
7	13	2.60	2.60	43.80
8	20	4.00	4.00	47.80
9	6	1.20	1.20	49.00
10	99	20.00	20.00	69.00
11	2	0.40	0.40	69.40
12	3	0.60	0.60	70.00
14	3	0.60	0.60	70.60
15	36	7.30	7.30	77.80
16	2	0.40	0.40	78.20
17	2	0.40	0.40	78.60
18	7	1.40	1.40	80.00
19	1	0.20	0.20	80.20
20	55	11.10	11.10	91.30
21	3	0.60	0.60	91.90
25	3	0.60	0.60	92.50
28	3	0.60	0.60	93.10
30	14	2.80	2.80	96.00
31	1	0.20	0.20	96.20
35	1	0.20	0.20	96.40
40	18	3.60	3.60	100.00
<b>Total</b>	496	100.00	100.00	

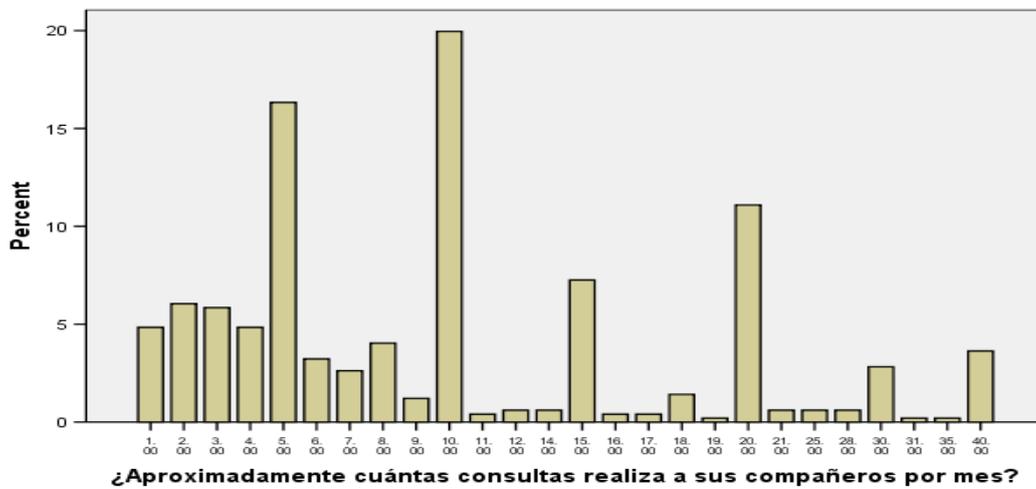
**Fuente:** Elaboración propia a través de Herramienta SPSS

**Tabla 16.** Operaciones Estadísticas ¿Cuántas consultas a través de TIC's realiza a sus compañeros por mes?

<b>Número de registros</b>	Validos	496
	Desaparecidos	0
<b>Media</b>		10.97
<b>Mediana</b>		10.00
<b>Moda</b>		10.00

**Fuente:** Elaboración propia a través de Herramienta SPSS

**¿Aproximadamente cuántas consultas realiza a sus compañeros por mes?**



**Figura 16.** Grafica de distribución de ¿Cuántas consultas a través de TIC's realiza a sus compañeros por mes?

**Fuente:** Generado desde herramienta SPSS

Aproximadamente cuántos recursos educativos descarga de la plataforma virtual cada mes?

**Tabla 17.** Frecuencias de ¿Cuántos recursos educativos descarga de la plataforma virtual cada mes?

	Frecuencia	Porcentaje	Porcentaje Valido	Porcentaje Acumulativo
1	31	6.30	6.30	6.30
2	52	10.50	10.50	16.70
3	35	7.10	7.10	23.80
4	30	6.00	6.00	29.80
5	55	11.10	11.10	40.90
6	12	2.40	2.40	43.30
7	45	9.10	9.10	52.40
8	18	3.60	3.60	56.00
9	6	1.20	1.20	57.30
10	78	15.70	15.70	73.00
11	2	0.40	0.40	73.40
12	12	2.40	2.40	75.80
15	29	5.80	5.80	81.70
16	3	0.60	0.60	82.30
17	2	0.40	0.40	82.70
18	2	0.40	0.40	83.10
19	1	0.20	0.20	83.30
20	43	8.70	8.70	91.90
21	2	0.40	0.40	92.30
24	1	0.20	0.20	92.50
25	5	1.00	1.00	93.50
30	15	3.00	3.00	96.60
40	17	3.40	3.40	100.00
Total	496	100.00	100.00	

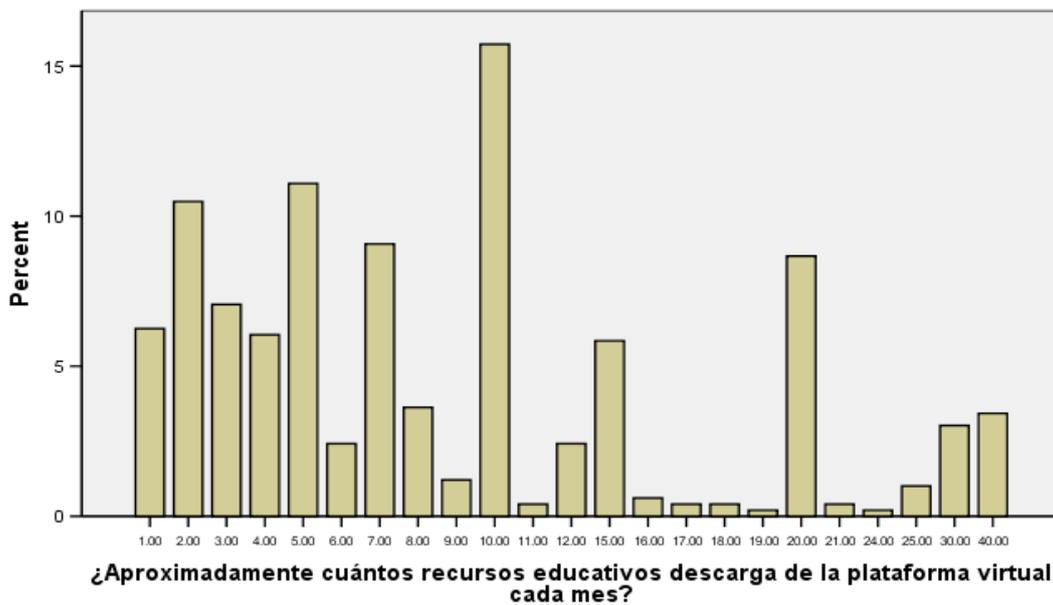
**Fuente:** Elaboración propia a través de Herramienta SPSS

**Tabla 18.** Operaciones Estadísticas ¿Cuántos recursos educativos descarga de la plataforma virtual cada mes?

<b>Número de registros</b>	Validos	496
	Desaparecidos	0
<b>Media</b>		9.92
<b>Mediana</b>		7.00
<b>Moda</b>		10.00

**Fuente:** Elaboración propia a través de Herramienta SPSS

**¿Aproximadamente cuántos recursos educativos descarga de la plataforma virtual cada mes?**



**Figura 17.** Grafica de distribución de ¿Cuántos recursos educativos descarga de la plataforma virtual cada mes?

**Fuente:** Generado desde herramienta SPSS

¿Aproximadamente cuántos videos académicos mira en youtube cada mes?

**Tabla 19.** Frecuencias de ¿Cuántos videos académicos mira en youtube cada mes?

	Frecuencia	Porcentaje	Porcentaje Valido	Porcentaje Acumulativo
1	38	7.70	7.70	7.70
2	51	10.30	10.30	17.90
3	50	10.10	10.10	28.00
4	33	6.70	6.70	34.70
5	93	18.80	18.80	53.40
6	15	3.00	3.00	56.50
7	12	2.40	2.40	58.90
8	18	3.60	3.60	62.50
9	6	1.20	1.20	63.70
10	79	15.90	15.90	79.60
12	1	0.20	0.20	79.80
13	2	0.40	0.40	80.20
14	1	0.20	0.20	80.40
15	20	4.00	4.00	84.50
16	1	0.20	0.20	84.70
18	1	0.20	0.20	84.90
19	1	0.20	0.20	85.10
20	36	7.30	7.30	92.30
24	1	0.20	0.20	92.50
25	4	0.80	0.80	93.30
30	33	6.70	6.70	100.00
<b>Total</b>	496	100.00	100.00	

**Fuente:** Elaboración propia a través de Herramienta SPSS

**Tabla 20.** Operaciones Estadísticas ¿Cuántos videos

académicos mira en youtube cada mes?

Número de registros	Validos	496
	Desaparecidos	0
Media		8.65
Mediana		5.00
Moda		5.00

Fuente: Elaboración propia a través de Herramienta SPSS

¿Aproximadamente cuántos videos académicos mira en youtube cada mes?

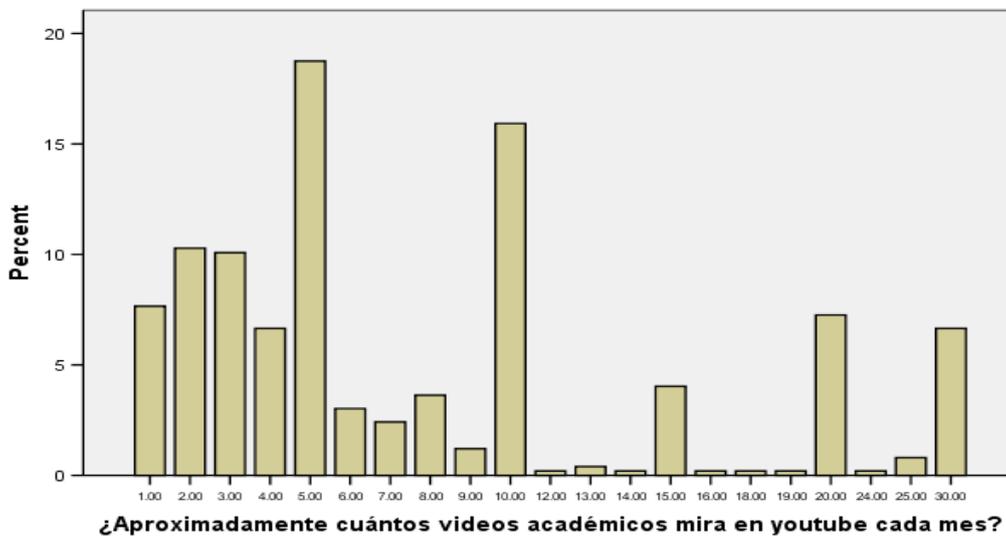


Figura 18. Grafica de distribución de ¿Cuántos videos académicos mira en youtube cada mes?

Fuente: Generado desde herramienta SPSS

¿Aproximadamente en cuántos foros virtuales participa cada mes?

**Tabla 21.** Frecuencias de ¿En cuántos foros virtuales participa cada mes?

	Frecuencia	Porcentaje	Porcentaje Valido	Porcentaje Acumulativo
0	138	27.80	27.80	27.80
1	85	17.10	17.10	45.00
2	104	21.00	21.00	65.90
3	41	8.30	8.30	74.20
4	19	3.80	3.80	78.00
5	58	11.70	11.70	89.70
6	4	0.80	0.80	90.50
7	4	0.80	0.80	91.30
8	4	0.80	0.80	92.10
9	2	0.40	0.40	92.50
10	18	3.60	3.60	96.20
11	1	0.20	0.20	96.40
12	1	0.20	0.20	96.60
14	1	0.20	0.20	96.80
15	16	3.20	3.20	100.00
Total	496	100.00	100.00	

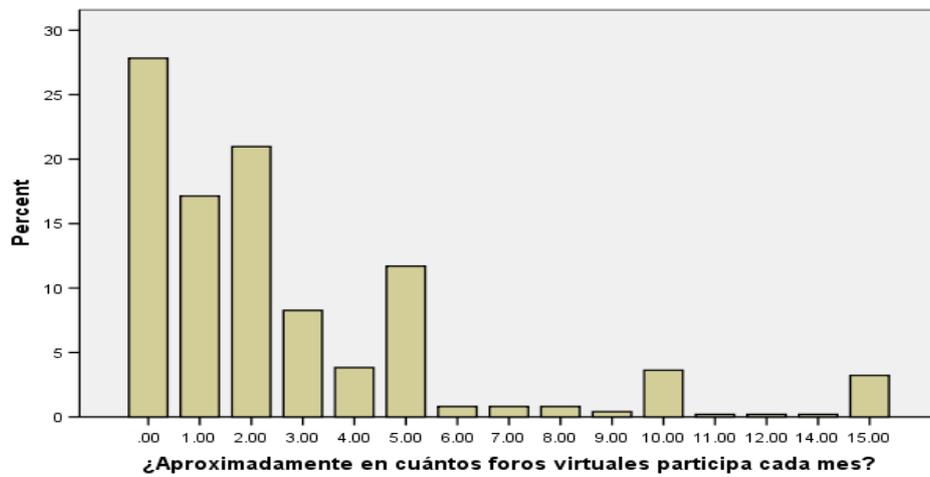
**Fuente:** Elaboración propia a través de Herramienta SPSS

**Tabla 22.** Operaciones Estadísticas ¿En cuántos foros virtuales participa cada mes?

<b>Número de registros</b>	Validos	496
	Desaparecidos	0
<b>Media</b>		2.70
<b>Mediana</b>		2.00
<b>Moda</b>		0.00

**Fuente:** Elaboración propia a través de Herramienta SPSS

¿Aproximadamente en cuántos foros virtuales participa cada mes?



**Figura 19.** Grafica de distribución de ¿En cuántos foros virtuales participa cada mes?

**Fuente:** Generado desde herramienta SPSS

¿Aproximadamente cuántos post o tweets sobre temas académicos realiza en las redes sociales por mes?

**Tabla 23.** Frecuencias de ¿Cuántos post o tweets sobre temas académicos realiza en las redes sociales por mes?

	Frecuencia	Porcentaje	Porcentaje Valido	Porcentaje Acumulativo
0	179	36.10	36.10	36.10
1	68	13.70	13.70	49.80
2	63	12.70	12.70	62.50
3	28	5.60	5.60	68.10
4	25	5.00	5.00	73.20
5	58	11.70	11.70	84.90
6	8	1.60	1.60	86.50
7	10	2.00	2.00	88.50
8	5	1.00	1.00	89.50
9	3	0.60	0.60	90.10
10	27	5.40	5.40	95.60
11	1	0.20	0.20	95.80
13	1	0.20	0.20	96.00
15	2	0.40	0.40	96.40
20	18	3.60	3.60	100.00
Total	496	100.00	100.00	

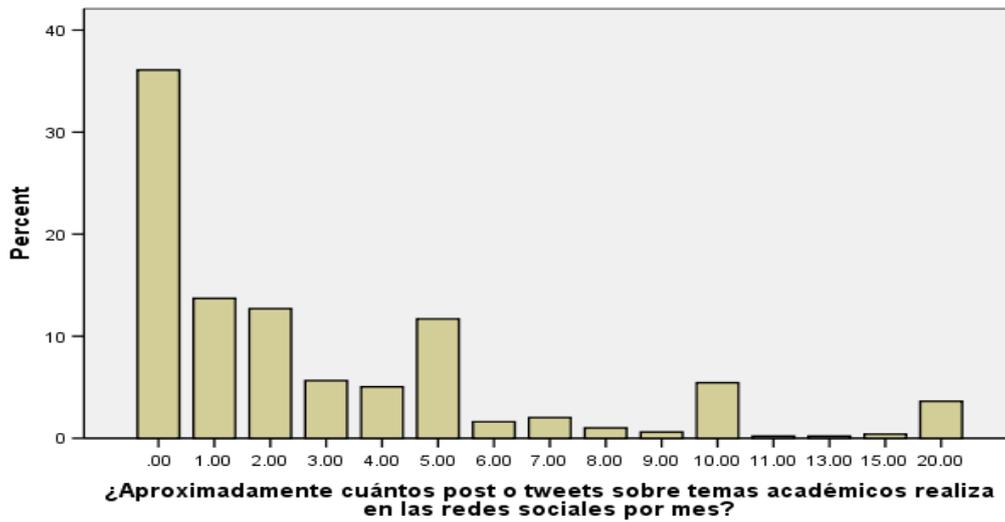
**Fuente:** Elaboración propia a través de Herramienta SPSS

**Tabla 24.** Operaciones Estadísticas ¿Cuántos post o tweets sobre temas académicos realiza en las redes sociales por mes?

Número de registros	Validos	496
	Desaparecidos	0
Media		3.10
Mediana		2.00
Moda		0.00

**Fuente:** Elaboración propia a través de Herramienta SPSS

**¿Aproximadamente cuántos post o tweets sobre temas académicos realiza en las redes sociales por mes?**



**Figura 20.** Grafica de distribución de ¿Cuántos post o tweets sobre temas académicos realiza en las redes sociales por mes?

**Fuente:** Generado desde herramienta SPSS

¿Aproximadamente cuántas horas chatea sobre temas académicos cada mes?

**Tabla 25.** Frecuencias de ¿Cuántas horas chatea sobre temas académicos cada mes?

	Frecuencia	Porcentaje	Porcentaje Valido	Porcentaje Acumulativo
0	33	6.70	6.70	6.70
1	50	10.10	10.10	16.70
2	60	12.10	12.10	28.80
3	37	7.50	7.50	36.30
4	17	3.40	3.40	39.70
5	49	9.90	9.90	49.60
6	13	2.60	2.60	52.20
7	12	2.40	2.40	54.60
8	17	3.40	3.40	58.10
9	2	0.40	0.40	58.50
10	74	14.90	14.90	73.40
12	6	1.20	1.20	74.60
13	3	0.60	0.60	75.20
14	1	0.20	0.20	75.40
15	22	4.40	4.40	79.80
16	1	0.20	0.20	80.00
18	4	0.80	0.80	80.80
19	1	0.20	0.20	81.00
20	37	7.50	7.50	88.50
21	1	0.20	0.20	88.70
24	3	0.60	0.60	89.30
25	2	0.40	0.40	89.70
28	2	0.40	0.40	90.10
29	2	0.40	0.40	90.50
30	18	3.60	3.60	94.20
32	1	0.20	0.20	94.40
35	1	0.20	0.20	94.60
40	5	1.00	1.00	95.60
42	1	0.20	0.20	95.80
50	21	4.20	4.20	100.00
Total	496	100.00	100.00	

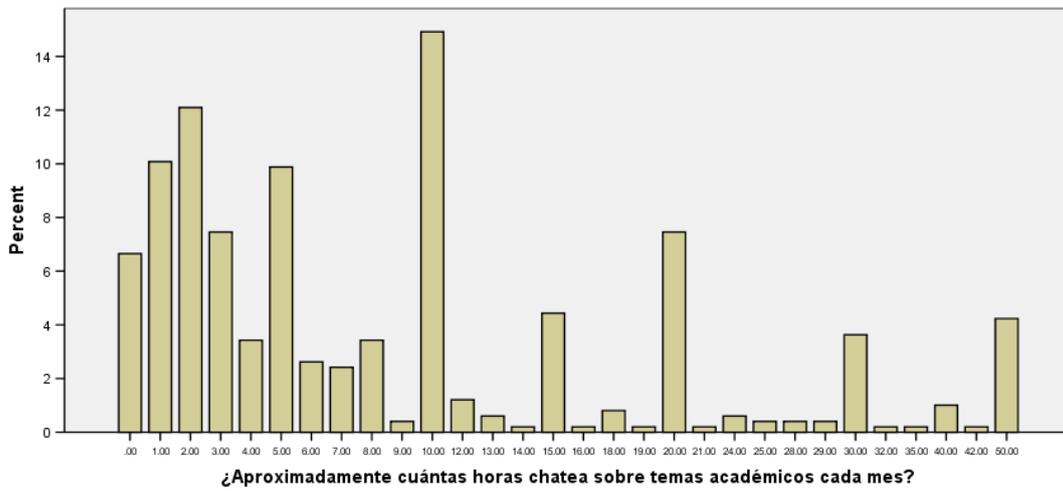
**Fuente:** Elaboración propia a través de Herramienta SPSS

**Tabla 26.** Operaciones Estadísticas ¿Cuántas horas chatea sobre temas académicos cada mes?

<b>Número de registros</b>	Validos	496
	Desaparecidos	0
<b>Media</b>		10.30
<b>Mediana</b>		6.00
<b>Moda</b>		10.00

**Fuente:** Elaboración propia a través de Herramienta SPSS

**¿Aproximadamente cuántas horas chatea sobre temas académicos cada mes?**



**Figura 21.** Gráfica de distribución de ¿Cuántas horas chatea sobre temas académicos cada mes?

**Fuente:** Generado desde herramienta SPSS

¿Aproximadamente cuántas horas busca información académica en internet cada mes?

**Tabla 27.** Frecuencias de ¿Cuántas horas busca información académica en internet cada mes?

	Frecuencia	Porcentaje	Porcentaje Valido	Porcentaje Acumulativo
0	2	0.40	0.40	0.40
1	18	3.60	3.60	4.00
2	26	5.20	5.20	9.30
3	26	5.20	5.20	14.50
4	16	3.20	3.20	17.70
5	29	5.80	5.80	23.60
6	14	2.80	2.80	26.40
7	10	2.00	2.00	28.40
8	16	3.20	3.20	31.70
9	3	0.60	0.60	32.30
10	53	10.70	10.70	42.90
11	1	0.20	0.20	43.10
12	9	1.80	1.80	45.00
13	2	0.40	0.40	45.40
14	5	1.00	1.00	46.40
15	28	5.60	5.60	52.00
16	1	0.20	0.20	52.20
18	3	0.60	0.60	52.80
19	2	0.40	0.40	53.20
20	57	11.50	11.50	64.70
21	1	0.20	0.20	64.90
22	1	0.20	0.20	65.10
24	6	1.20	1.20	66.30
25	13	2.60	2.60	69.00
26	1	0.20	0.20	69.20
30	41	8.30	8.30	77.40
31	2	0.40	0.40	77.80
32	3	0.60	0.60	78.40
34	1	0.20	0.20	78.60
36	2	0.40	0.40	79.00
40	20	4.00	4.00	83.10
45	1	0.20	0.20	83.30

46	1	0.20	0.20	83.50
48	4	0.80	0.80	84.30
50	19	3.80	3.80	88.10
60	11	2.20	2.20	90.30
70	1	0.20	0.20	90.50
72	1	0.20	0.20	90.70
75	1	0.20	0.20	90.90
80	12	2.40	2.40	93.30
96	1	0.20	0.20	93.50
100	10	2.00	2.00	95.60
120	22	4.40	4.40	100.00
Total	496	100.00	100.00	

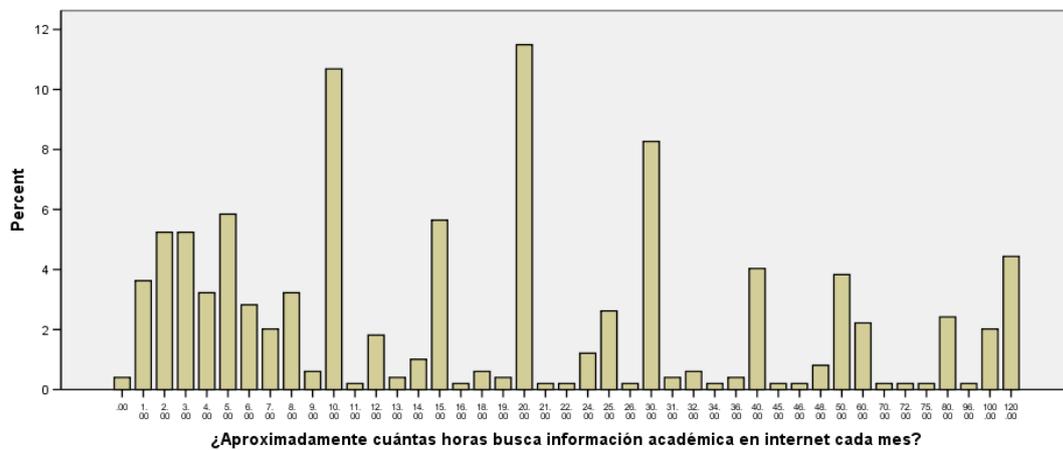
Fuente: Elaboración propia a través de Herramienta SPSS

**Tabla 28.** Operaciones Estadísticas ¿Cuántas horas busca información académica en internet cada mes?

Número de registros	Validos	496
	Desaparecidos	0
Media		25.64
Mediana		15.00
Moda		20.00

Fuente: Elaboración propia a través de Herramienta SPSS

¿Aproximadamente cuántas horas busca información académica en internet cada mes?



**Figura 22.** Grafica de distribución de ¿Cuántas horas busca información académica en internet cada mes?

Fuente: Generado desde herramienta SPSS

¿Aproximadamente cuántas horas utiliza la biblioteca virtual de la universidad cada mes?

**Tabla 29.** Frecuencias de ¿Cuántas horas utiliza la biblioteca virtual de la universidad cada mes?

	Frecuencia	Porcentaje	Porcentaje Valido	Porcentaje Acumulativo
1	45	9.10	9.10	9.10
2	57	11.50	11.50	20.60
3	29	5.80	5.80	26.40
4	130	26.20	26.20	52.60
5	57	11.50	11.50	64.10
6	13	2.60	2.60	66.70
7	11	2.20	2.20	69.00
8	12	2.40	2.40	71.40
9	2	0.40	0.40	71.80
10	46	9.30	9.30	81.00
11	4	0.80	0.80	81.90
12	9	1.80	1.80	83.70
14	1	0.20	0.20	83.90
15	10	2.00	2.00	85.90
16	3	0.60	0.60	86.50
17	1	0.20	0.20	86.70
18	1	0.20	0.20	86.90
20	20	4.00	4.00	90.90
21	1	0.20	0.20	91.10
22	1	0.20	0.20	91.30
24	1	0.20	0.20	91.50
25	4	0.80	0.80	92.30
30	13	2.60	2.60	95.00
35	1	0.20	0.20	95.20
38	1	0.20	0.20	95.40
40	3	0.60	0.60	96.00
46	1	0.20	0.20	96.20
48	2	0.40	0.40	96.60
50	17	3.40	3.40	100.00
Total	496	100.00	100.00	

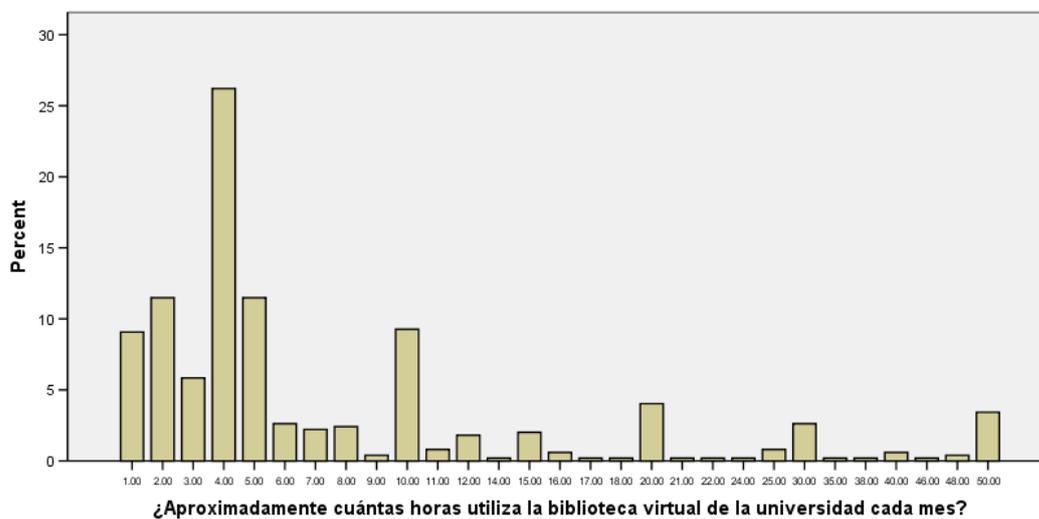
**Fuente:** Elaboración propia a través de Herramienta SPSS

**Tabla 30.** Operaciones Estadísticas ¿Cuántas horas utiliza la biblioteca virtual de la universidad cada mes?

<b>Número de registros</b>	Validos	496
	Desaparecidos	0
<b>Media</b>		8.71
<b>Mediana</b>		4.00
<b>Moda</b>		4.00

**Fuente:** Elaboración propia a través de Herramienta SPSS

**¿Aproximadamente cuántas horas utiliza la biblioteca virtual de la universidad cada mes?**



**Figura 23.** Grafica de distribución de ¿Cuántas horas utiliza la biblioteca virtual de la universidad cada mes?

**Fuente:** Generado desde herramienta SPSS

### 3. DataSet - Usos internet para actividades de entretenimiento.

¿Aproximadamente cuántas horas a la semana chatea por diversión?

**Tabla 31.** Frecuencias de ¿Cuántas horas a la semana chatea por diversión?

	Frecuencia	Porcentaje	Porcentaje Valido	Porcentaje Acumulativo
1	21	4.20	4.20	4.20
2	33	6.70	6.70	10.90
3	23	4.60	4.60	15.50
4	10	2.00	2.00	17.50
5	38	7.70	7.70	25.20
6	15	3.00	3.00	28.20
7	16	3.20	3.20	31.50
8	22	4.40	4.40	35.90
9	2	0.40	0.40	36.30
10	75	15.10	15.10	51.40
12	8	1.60	1.60	53.00
13	2	0.40	0.40	53.40
14	8	1.60	1.60	55.00
15	32	6.50	6.50	61.50
16	3	0.60	0.60	62.10
17	1	0.20	0.20	62.30
18	3	0.60	0.60	62.90
19	1	0.20	0.20	63.10
20	39	7.90	7.90	71.00
21	5	1.00	1.00	72.00
22	1	0.20	0.20	72.20
23	1	0.20	0.20	72.40
24	8	1.60	1.60	74.00
25	12	2.40	2.40	76.40
26	2	0.40	0.40	76.80
28	4	0.80	0.80	77.60
29	1	0.20	0.20	77.80
30	22	4.40	4.40	82.30
32	1	0.20	0.20	82.50
35	6	1.20	1.20	83.70
36	1	0.20	0.20	83.90

40	11	2.20	2.20	86.10
42	1	0.20	0.20	86.30
48	2	0.40	0.40	86.70
50	13	2.60	2.60	89.30
60	10	2.00	2.00	91.30
69	1	0.20	0.20	91.50
70	8	1.60	1.60	93.10
72	5	1.00	1.00	94.20
80	2	0.40	0.40	94.60
84	27	5.40	5.40	100.00
Total	496	100.00	100.00	

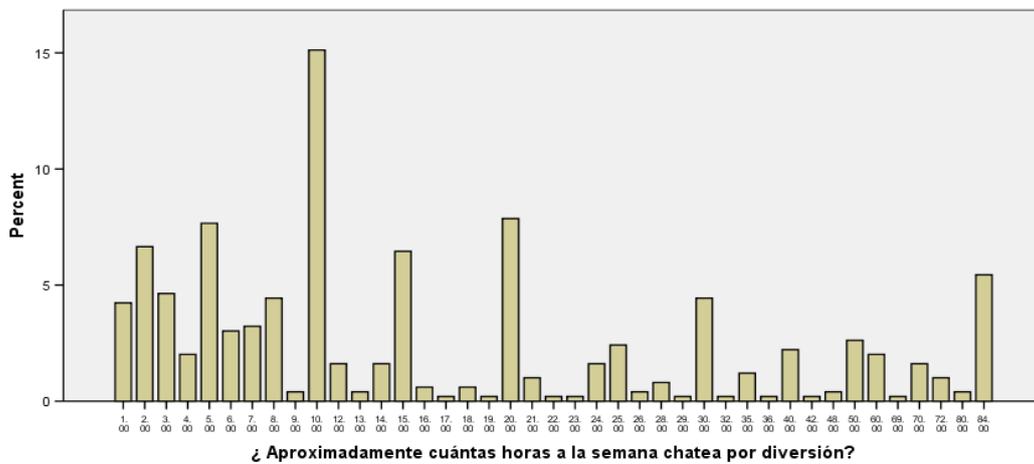
Fuente: Elaboración propia a través de Herramienta SPSS

**Tabla 32.** Operaciones Estadísticas ¿Cuántas horas a la semana chatea por diversión?

Número de registros	Validos	496
	Desaparecidos	0
Media		20.53
Mediana		10.00
Moda		10.00

Fuente: Elaboración propia a través de Herramienta SPSS

¿ Aproximadamente cuántas horas a la semana chatea por diversión?



**Figura 24.** Grafica de distribución de ¿Cuántas horas a la semana chatea por diversión?

Fuente: Generado desde herramienta SPSS

¿Aproximadamente cuántas horas a la semana utiliza redes sociales?

**Tabla 33.** Frecuencias de ¿Cuántas horas a la semana utiliza redes sociales?

	Frecuencia	Porcentaje	Porcentaje Valido	Porcentaje Acumulativo
1	13	2.60	2.60	2.60
2	16	3.20	3.20	5.80
3	25	5.00	5.00	10.90
4	14	2.80	2.80	13.70
5	42	8.50	8.50	22.20
6	13	2.60	2.60	24.80
7	18	3.60	3.60	28.40
8	17	3.40	3.40	31.90
9	5	1.00	1.00	32.90
10	58	11.70	11.70	44.60
12	10	2.00	2.00	46.60
13	2	0.40	0.40	47.00
14	7	1.40	1.40	48.40
15	35	7.10	7.10	55.40
16	4	0.80	0.80	56.30
17	2	0.40	0.40	56.70
18	3	0.60	0.60	57.30
19	2	0.40	0.40	57.70
20	58	11.70	11.70	69.40
21	2	0.40	0.40	69.80
24	7	1.40	1.40	71.20
25	15	3.00	3.00	74.20
27	1	0.20	0.20	74.40
28	3	0.60	0.60	75.00
30	22	4.40	4.40	79.40
32	1	0.20	0.20	79.60
33	1	0.20	0.20	79.80
35	9	1.80	1.80	81.70
38	1	0.20	0.20	81.90
40	14	2.80	2.80	84.70
42	3	0.60	0.60	85.30
48	2	0.40	0.40	85.70
49	1	0.20	0.20	85.90

50	13	2.60	2.60	88.50
60	9	1.80	1.80	90.30
70	8	1.60	1.60	91.90
72	4	0.80	0.80	92.70
75	1	0.20	0.20	92.90
78	1	0.20	0.20	93.10
80	8	1.60	1.60	94.80
84	1	0.20	0.20	95.00
90	25	5.00	5.00	100.00
Total	496	100.00	100.00	

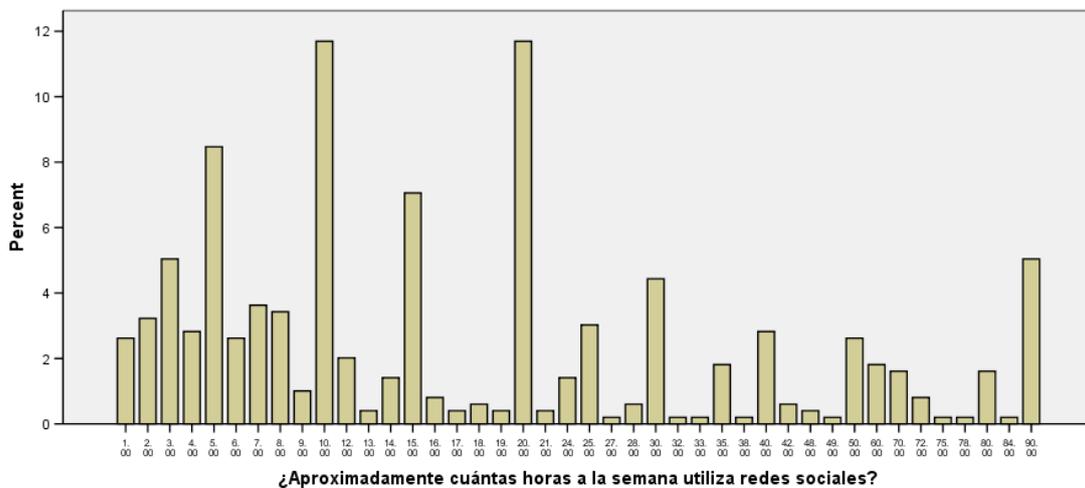
Fuente: Elaboración propia a través de Herramienta SPSS

**Tabla 34.** Operaciones Estadísticas ¿Cuántas horas a la semana utiliza redes sociales?

Número de registros	Validos	496
	Desaparecidos	0
Media		22.66
Mediana		15.00
Moda		10.00(a)

Fuente: Elaboración propia a través de Herramienta SPSS

¿Aproximadamente cuántas horas a la semana utiliza redes sociales?



**Figura 25.** Grafica de distribución de ¿Cuántas horas a la semana utiliza redes sociales?

Fuente: Generado desde herramienta SPSS

¿Aproximadamente cuántas horas a la semana utiliza juegos en línea?

**Tabla 35.** Frecuencias de ¿Cuántas horas a la semana utiliza juegos en línea?

	Frecuencia	Porcentaje	Porcentaje Valido	Porcentaje Acumulativo
0	254	51.20	51.20	51.20
1	38	7.70	7.70	58.90
2	45	9.10	9.10	67.90
3	10	2.00	2.00	70.00
4	14	2.80	2.80	72.80
5	30	6.00	6.00	78.80
6	5	1.00	1.00	79.80
7	2	0.40	0.40	80.20
8	5	1.00	1.00	81.30
9	1	0.20	0.20	81.50
10	35	7.10	7.10	88.50
12	5	1.00	1.00	89.50
14	1	0.20	0.20	89.70
15	6	1.20	1.20	90.90
18	2	0.40	0.40	91.30
19	1	0.20	0.20	91.50
20	15	3.00	3.00	94.60
24	2	0.40	0.40	95.00
28	2	0.40	0.40	95.40
30	7	1.40	1.40	96.80
40	6	1.20	1.20	98.00
45	1	0.20	0.20	98.20
50	9	1.80	1.80	100.00
Total	496	100.00	100.00	

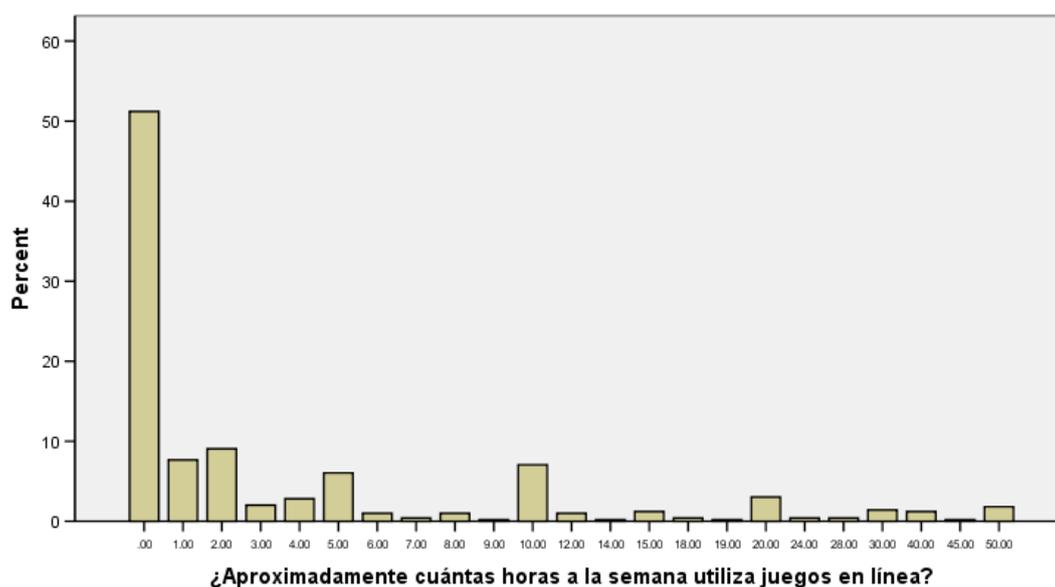
**Fuente:** Elaboración propia a través de Herramienta SPSS

**Tabla 36.** Operaciones Estadísticas ¿Cuántas horas a la semana utiliza juegos en línea?

<b>Número de registros</b>	Validos	496
	Desaparecidos	0
<b>Media</b>		4.79
<b>Mediana</b>		0.00
<b>Moda</b>		0.00

**Fuente:** Elaboración propia a través de Herramienta SPSS

**¿Aproximadamente cuántas horas a la semana utiliza juegos en línea?**



**Figura 26.** Grafica de distribución de ¿Cuántas horas a la semana utiliza juegos en línea?

**Fuente:** Generado desde herramienta SPSS

¿Aproximadamente cuántas horas a la semana descarga música, videos y programas?

**Tabla 37.** Frecuencias de ¿Cuántas horas a la semana descarga música, videos y programas?

	Frecuencia	Porcentaje	Porcentaje Valido	Porcentaje Acumulativo
1	47	9.50	9.50	9.50
2	77	15.50	15.50	25.00
3	40	8.10	8.10	33.10
4	18	3.60	3.60	36.70
5	110	22.20	22.20	58.90
6	9	1.80	1.80	60.70
7	14	2.80	2.80	63.50
8	11	2.20	2.20	65.70
9	3	0.60	0.60	66.30
10	57	11.50	11.50	77.80
11	1	0.20	0.20	78.00
12	4	0.80	0.80	78.80
14	1	0.20	0.20	79.00
15	13	2.60	2.60	81.70
16	1	0.20	0.20	81.90
19	1	0.20	0.20	82.10
20	31	6.30	6.30	88.30
21	1	0.20	0.20	88.50
24	5	1.00	1.00	89.50
25	9	1.80	1.80	91.30
30	15	3.00	3.00	94.40
35	1	0.20	0.20	94.60
38	1	0.20	0.20	94.80
40	2	0.40	0.40	95.20
42	1	0.20	0.20	95.40
48	1	0.20	0.20	95.60
50	7	1.40	1.40	97.00
60	15	3.00	3.00	100.00
Total	496	100.00	100.00	

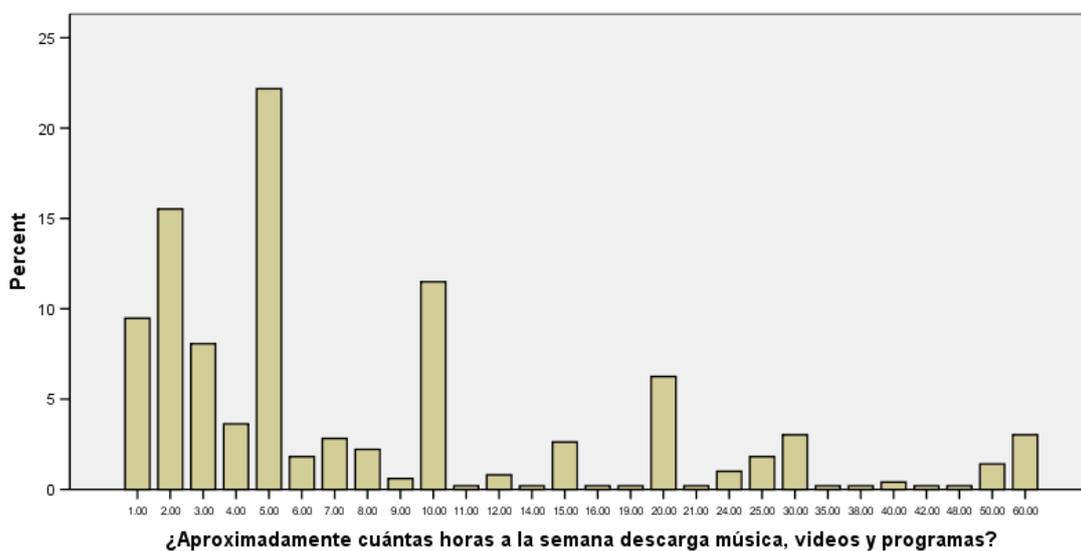
**Fuente:** Elaboración propia a través de Herramienta SPSS

**Tabla 38.** Operaciones Estadísticas ¿Cuántas horas a la semana descarga música, videos y programas?

<b>Número de registros</b>	Validos	496
	Desaparecidos	0
<b>Media</b>	10.10	
<b>Mediana</b>	5.00	
<b>Moda</b>	5.00	

**Fuente:** Elaboración propia a través de Herramienta SPSS

**¿Aproximadamente cuántas horas a la semana descarga música, videos y programas?**



**Figura 27.** Grafica de distribución de ¿Cuántas horas a la semana descarga música, videos y programas?

**Fuente:** Generado desde herramienta SPSS

¿Aproximadamente cuántos videos para entretenimiento mira en youtube cada semana?

**Tabla 39.** Frecuencias de ¿Cuántos videos para entretenimiento mira en youtube cada semana?

	Frecuencia	Porcentaje	Porcentaje Valido	Porcentaje Acumulativo
1	36	7.30	7.30	7.30
2	52	10.50	10.50	17.70
3	37	7.50	7.50	25.20
4	15	3.00	3.00	28.20
5	64	12.90	12.90	41.10
6	14	2.80	2.80	44.00
7	32	6.50	6.50	50.40
8	20	4.00	4.00	54.40
9	1	0.20	0.20	54.60
10	61	12.30	12.30	66.90
12	9	1.80	1.80	68.80
13	2	0.40	0.40	69.20
14	1	0.20	0.20	69.40
15	31	6.30	6.30	75.60
16	3	0.60	0.60	76.20
17	1	0.20	0.20	76.40
18	1	0.20	0.20	76.60
19	1	0.20	0.20	76.80
20	32	6.50	6.50	83.30
24	6	1.20	1.20	84.50
25	9	1.80	1.80	86.30
28	3	0.60	0.60	86.90
29	1	0.20	0.20	87.10
30	19	3.80	3.80	90.90
32	1	0.20	0.20	91.10
35	1	0.20	0.20	91.30
40	4	0.80	0.80	92.10
45	1	0.20	0.20	92.30
48	1	0.20	0.20	92.50
50	7	1.40	1.40	94.00
58	1	0.20	0.20	94.20

60	5	1.00	1.00	95.20
70	1	0.20	0.20	95.40
80	23	4.60	4.60	100.00
Total	496	100.00	100.00	

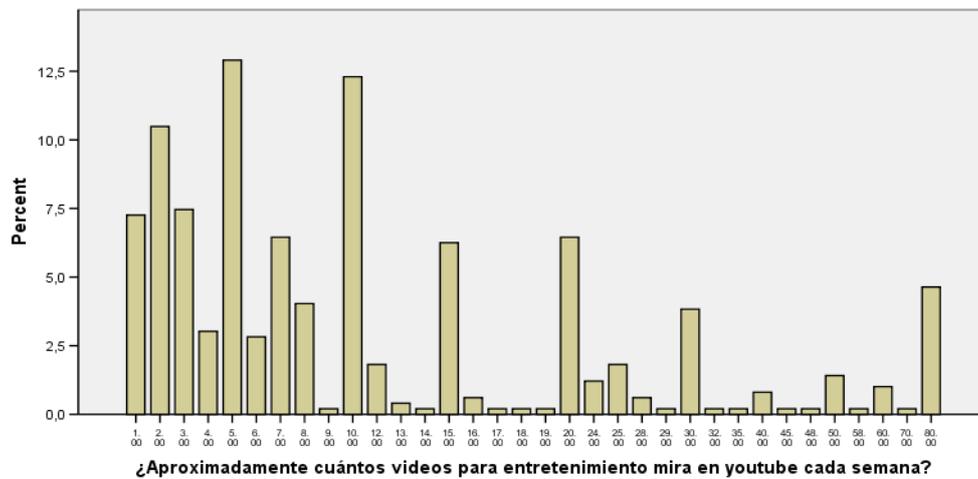
**Fuente:** Elaboración propia a través de Herramienta SPSS

**Tabla 40.** Operaciones Estadísticas ¿Cuántos videos para entretenimiento mira en youtube cada semana?

<b>Número de registros</b>	Validos	496
	Desaparecidos	0
<b>Media</b>		14.24
<b>Mediana</b>		7.00
<b>Moda</b>		5.00

**Fuente:** Elaboración propia a través de Herramienta SPSS

¿Aproximadamente cuántos videos para entretenimiento mira en youtube cada semana?



**Figura 28.** Grafica de distribución de ¿Cuántos videos para entretenimiento mira en youtube cada semana?

**Fuente:** Generado desde herramienta SPSS

#### 4. DataSet – Rendimiento académico.

En el semestre anterior, ¿en cuántas asignaturas se matriculó?

**Tabla 41.** Frecuencias de ¿En cuántas asignaturas se matriculó?

	Frecuencia	Porcentaje	Porcentaje Valido	Porcentaje Acumulativo
0	1	0.20	0.20	0.20
1	3	0.60	0.60	0.80
2	2	0.40	0.40	1.20
3	6	1.20	1.20	2.40
4	11	2.20	2.20	4.60
5	79	15.90	15.90	20.60
6	98	19.80	19.80	40.30
7	137	27.60	27.60	67.90
8	79	15.90	15.90	83.90
9	31	6.30	6.30	90.10
10	9	1.80	1.80	91.90
11	34	6.90	6.90	98.80
12	3	0.60	0.60	99.40
13	3	0.60	0.60	100.00
Total	496	100.00	100.00	

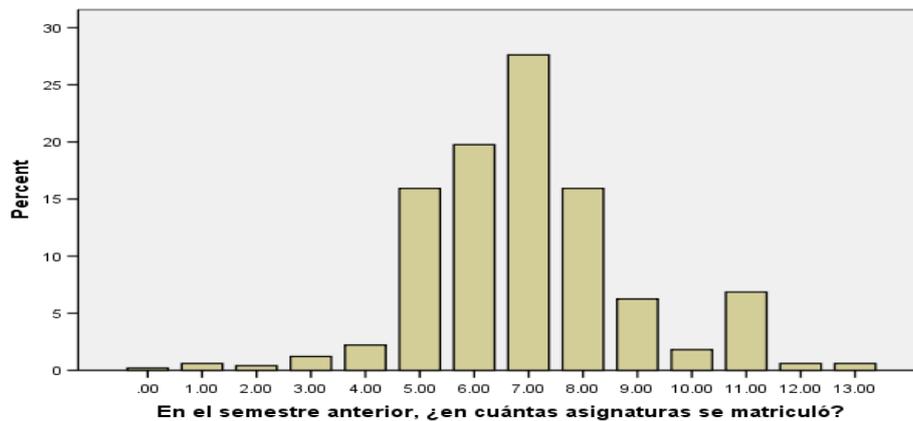
**Fuente:** Elaboración propia a través de Herramienta SPSS

**Tabla 42.** Operaciones Estadísticas ¿En cuántas asignaturas se matriculó?

Número de registros	Validos	496
	Desaparecidos	0
Media		6.98
Mediana		7.00
Moda		7.00

**Fuente:** Elaboración propia a través de Herramienta SPSS

En el semestre anterior, ¿en cuántas asignaturas se matriculó?



**Figura 29.** Grafica de distribución de ¿En cuántas asignaturas se matriculó?

**Fuente:** Generado desde herramienta SPSS

En el semestre anterior ¿cuántas asignaturas aprobó?

**Tabla 43.** Frecuencias de ¿Cuántas asignaturas aprobó?

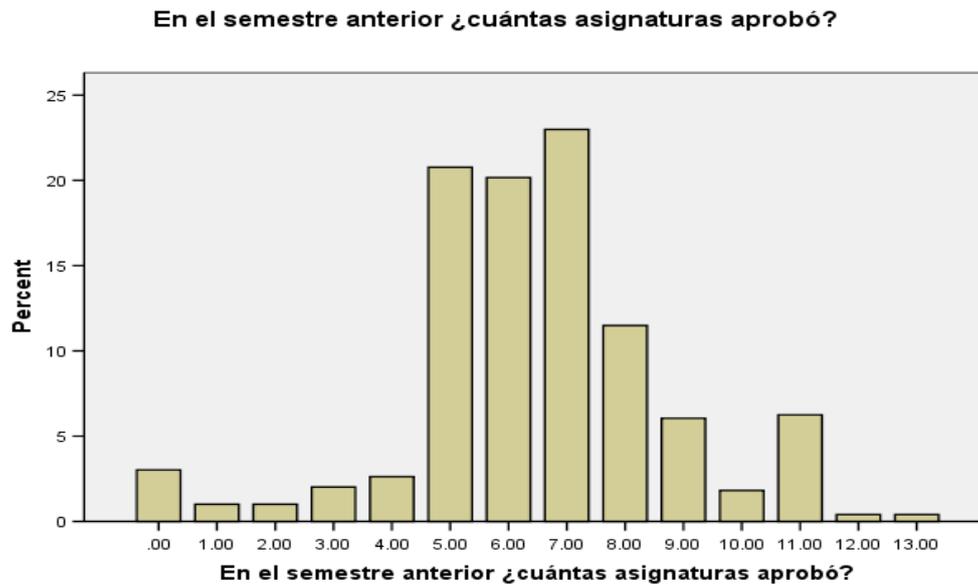
	Frecuencia	Porcentaje	Porcentaje Valido	Porcentaje Acumulativo
0	15	3.00	3.00	3.00
1	5	1.00	1.00	4.00
2	5	1.00	1.00	5.00
3	10	2.00	2.00	7.10
4	13	2.60	2.60	9.70
5	103	20.80	20.80	30.40
6	100	20.20	20.20	50.60
7	114	23.00	23.00	73.60
8	57	11.50	11.50	85.10
9	30	6.00	6.00	91.10
10	9	1.80	1.80	92.90
11	31	6.30	6.30	99.20
12	2	0.40	0.40	99.60
13	2	0.40	0.40	100.00
Total	496	100.00	100.00	

**Fuente:** Elaboración propia a través de Herramienta SPSS

**Tabla 44.** Operaciones Estadísticas ¿Cuántas asignaturas aprobó?

Número de registros	Validos	496
	Desaparecidos	0
Media		6.49
Mediana		6.00
Moda		7.00

**Fuente:** Elaboración propia a través de Herramienta SPSS



**Figura 30.** Grafica de distribución de ¿Cuántas asignaturas aprobó?

**Fuente:** Generado desde herramienta SPSS

#### 6.2.4. Verificación de calidad de los datos

Luego de haber culminado la exploración de los datos que forman parte de los DataSet, se ha evidenciado que la mayoría de los mismos registran una buena calidad de datos, sin embargo en algunos campos se encontraron novedades de calidad, mismas que se detallan a continuación:

- **as\_mat\_se** .- Este campo contiene información respecto a “En el semestre anterior, ¿en cuántas asignaturas se matriculó?” y en el mismo existe un registro con un dato en cero (0).
- **as\_apr\_se** .- Este campo contiene información respecto a “En el semestre anterior ¿cuántas asignaturas aprobó?” y en el mismo existen quince registros con un dato en cero (0).

Adicionalmente es necesario indicar que al momento del traspaso de información física de las cuentas a digital, se procedió a excluir cuatro encuestas por tener un número significativo de campos vacíos.

La calidad de los datos se analizará, definirá y se depurará completamente en la Fase III de este proceso.

### **6.3. Fase III – Preparación de los datos**

Esta fase del proceso se iniciará seleccionando los DataSet definitivos a partir de los atributos y datos recogidos y explorados en la fase anterior; para luego proceder a su preparación para adaptarlos a las técnicas de Minería de Datos a ser aplicadas más adelante; esta preparación de datos incluye tareas de limpieza y transformación de tablas, registros y atributos con el fin de lograr calidad en la información resultante.

#### **6.3.1. Selección de Datos**

Como ya se lo había mencionado y descrito de manera preliminar en la fase anterior, para la aplicación de las técnicas de minería de datos se ha definido tres DataSet que van acorde a las hipótesis que se necesita confirmar, mismas que son:

✓ *DataSet – Ingreso Económico del estudiante*

Contiene los datos de ingreso familiar del estudiante, está compuesto de un campo **ing**, mismo que podrá tener registrado uno de cinco rangos de acuerdo al dato ingresado en la encuesta. Este DataSet será relacionado con los DataSet *Uso internet en actividades de académicas* y *Uso internet en actividades de entretenimiento*, para la generación del modelo.

✓ *DataSet – Uso internet en actividades académicas*

Contiene los datos de las diferentes formas de uso del internet en actividades académicas por parte del estudiante, este DataSet está compuesto de los campos detallados en la [Sección 6.2, punto 6.2.2, tabla 4], y será relacionado con los DataSet *Ingreso económico del estudiante* y *Rendimiento académico*, para la generación del modelo.

✓ *DataSet – Uso internet en actividades de entretenimiento.*

Contiene los datos de las diferentes formas de uso del internet en actividades de entretenimiento por parte del estudiante, está compuesto de los campos detallados en la [Sección 6.2, punto 6.2.2, tabla 5], y será relacionado con dos

DataSet: *Ingreso económico del estudiante y Rendimiento académico*, para la generación del modelo.

✓ *Rendimiento académico*

Contiene los datos de número de materias en las que se matriculo el estudiante encuestado, y el número de materias que aprobó, está compuesto de los campos detallados en la [Sección 6.2, punto 6.2.2, tabla 6] y será relacionado con los DataSet *Uso internet en actividades de académicas* y *Uso internet en actividades de entretenimiento*, para la generación del modelo.

### 6.3.2. Limpieza de los Datos

Se ha procedido a realizar la limpieza de los datos de acuerdo a lo levantado en [Sección 6.2, punto 0] de verificación de calidad de datos; esta limpieza consistió en establecer una definición que dé solución a los registros que en la información levantada se encontraban vacíos o con valores errados, para luego proceder a ejecutar la limpieza de los datos de la siguiente manera:

- ✓ Atributo **as\_mat\_se** .- En este atributo existe un registro con dato cero (0), para lo cual se procedió a reemplazar el dato de este registro con el valor de la *media* resultante respecto a todos los registros que conforman el atributo.
- ✓ Atributo **as\_apr\_se** .- En este atributo existen quince registros con dato cero (0), para lo cual se procedió a reemplazar los datos de estos registros con el valor de la *media* resultante respecto a todos los registros que conforman el atributo.

### 6.3.3. Estructuración e Integración de los Datos

En esta fase nos enfocamos a la verificación y generación de nuevos atributos a partir de otros ya existentes, para facilidad de adaptación de los datos a las técnicas de Minería de datos que aplicaríamos más adelante.

Dentro de los DataSet definidos tenemos el de *Rendimiento académico*, el cual estaba originalmente compuesto por dos atributos, sin embargo con el fin de hacer posible la aplicabilidad de las técnicas de minería de datos, se vio necesaria la creación de un

nuevo atributo como resultado de la diferencia entre atributo *as\_mat\_se* y el atributo *as\_apr\_se*, mismo que se tipificará como *as\_rep\_se* (*asignaturas reprobadas*), que presentaría el resultado de la resta entre las materias en las que se matriculo menos las aprobadas. Para lograr lo antes indicado se procedió a aplicar el filtro *AddExpression* de la herramienta Weka, el mismo que permitió agregar al final un atributo resultante de una función, es necesario especificar la fórmula que describe este atributo, en donde podemos calcular dicho atributo a partir de los valores de otro u otros ya existentes. Para mayor detalle del proceso aplicado y de los parámetros ingresados para lograr la creación del nuevo atributo diríjase al [ANEXO 3]. Las operaciones estadísticas en el nuevo atributo están descritas en [Figura 31].

### **Función**

$$a17 - a18 = a19$$

Selected attribute	
Name: as_rep_se	Type: Numeric
Missing: 0 (0%)	Distinct: 6
	Unique: 0 (0%)
Statistic	Value
Minimum	0
Maximum	6
Mean	0.373
StdDev	0.895

**Figura 31.** Operaciones estadísticas como resultado del nuevo atributo

**Fuente:** Generado desde herramienta WEKA

Luego de ello, con la finalidad de hacer más fácil la aplicación de la relación del “rendimiento académico” con los otros DataSet, se procedió a crear atributo categórico de cuatro niveles con nombre *rendimiento\_academico*, tomando como base el atributo *as\_rep\_se*. En este nuevo atributo categórico el nivel 1 figura a los estudiantes que no reprobaron asignaturas, el nivel 2 figura a los estudiantes que reprobaron una asignatura, el nivel 3 figura a los estudiantes que reprobaron dos asignaturas, y el nivel 4 figura a los estudiantes que reprobaron tres o más asignaturas. Para mayor detalle

del proceso aplicado y de los parámetros ingresados para logra la creación del nuevo atributo diríjase al [ANEXO 4].

#### **6.3.4. Formateo de los Datos**

Se ha procedido a cambiar al formato necesario a cada uno de los campos que conforman los cuatro DataSet, esto para hacer factible la aplicación de las técnicas de minería de datos. No se procede a depurar comas, tabuladores, caracteres especiales en los datos en este punto del proceso ya que los mismos fueron depurados al momento de digitalización de información.

#### **6.4. Fase IV – Modelado**

En esta fase de CRISP-DM se procede a seleccionar las técnicas y a construir el modelo de DM que nos ayudará a descubrir las hipótesis planteadas para esta investigación.

##### **6.4.1. Selección de la técnica de modelado**

Luego de haber definido los DataSet, y haber depurado inconsistencias mediante el proceso de limpieza de datos, se procede con la creación del modelo de minería de datos, seleccionando las técnicas más apropiadas para lograr dar solución al problema, dichas técnicas fueron elegidas en función a los siguientes criterios sugeridos por (J. A. Gallardo, n.d.):

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

Las técnicas y algoritmos que se ha procedido a aplicar para la construcción de este modelo de DM se detallan en la [Tabla 7].

**Tabla 45.** Técnicas y Algoritmos DM

TECNICA DM	ALGORITMO APLICADO
Análisis Factorial	CfsSubsetEval
Clasificación (Comprobación)	DecisionTable
Cluster	EM
Cluster	Simple K-means
Regresión Logística	Logistic
Métodos de reproceso	Add
Métodos de reproceso	AddExpression
Métodos de reproceso	AddCluster

**Fuente:** Elaboración propia

#### 6.4.2. Construcción del modelo

Para la construcción del modelo se ha procedió a aplicar técnicas de Minería de Datos de donde se obtuvieron los resultados explicados a continuación.

##### 6.4.2.1. Análisis Factorial (Reducción de variables)

Para realizar el análisis factorial se ha procedido a utilizar el algoritmo *CfsSubsetEval* pues el mismo permite la reducción de datos mediante la correlación de un conjunto de variables dadas, obteniendo como resultado un menor número de variables nuevas. Más detalles respecto a la presente técnica se encuentran en [Sección 5.5, punto 5.5.1].

Adicionalmente se evidenció la necesidad de aplicar el algoritmo de filtro *PartitionedMultifilter* mismo que permite la aplicación de filtros en subconjuntos de atributos definidos a nivel de parámetro, y el montaje de la salida se da en un nuevo conjunto de datos. Los atributos que no están cubiertos por cualquiera de los rangos pueden ser retenidos o retirados de la salida de acuerdo a como se lo haya definido.

Para realizar la aplicación de este proceso de simplificación de atributos en donde se inició utilizando el algoritmo *PartitionedMultifilter* se necesitó realizar algunas configuraciones solicitadas por la herramienta Weka.

El primer parámetro llamado *debug* permite activar la salida de información de depuración, por lo que al no necesitar esto se dejó por defecto en “false”; el segundo campo llamado *filters* permite seleccionar el algoritmo base con el cual se va a realizar el modelo, por lo que para este caso al requerir la aplicación de la técnica “análisis factorial” se parametrizó el algoritmo *CfsSubsetEval*, el tercer parámetro llamado *ranges* que otorga la capacidad de indicar los rango de atributos que se van a tomar para aplicar el algoritmo escogido en el parámetro dos, por lo que se registró dos rangos de atributos: del 1-10 respecto a atributos correspondientes a “uso de internet en actividades académicas” y del 11-15 respecto a atributos correspondientes a “uso de internet en actividades de entretenimiento”; y finalmente el parámetro *removeUnised* permite que a los atributos que no están siendo utilizados dentro de los rangos parametrizados se eliminen o se conserven, por lo que para este ejercicio se dejaron por default en “false” para mantener los atributos que no están participando en la aplicación de este algoritmo.

En el párrafo anterior se mencionó al algoritmo *CfsSubsetEval* como un parámetro a ingresar, sin embargo este algoritmo es quien permite la aplicación de la técnica de análisis factorial o simplificación de atributos, pues este evalúa el valor de un subconjunto de atributos considerando la capacidad de predicción individual de cada característica, junto con el grado de redundancia entre ellos, se prefieren los subconjuntos de características que están altamente correlacionados con la clase, mientras que tiene una baja intercorrelación. Para la correcta aplicación de este algoritmo evaluador, el mismo debe ir acompañado de un algoritmo de búsqueda, para lo cual para este ejercicio se ha definido utilizar *BestFirst*, el cual permite buscar interacciones entre atributos; este método va analizando lo que mejora y empeora un grupo de atributos al añadir elementos, con la posibilidad de hacer retrocesos para explorar con más detalle. Los parámetros realizados para este proceso se pueden encontrar en [ANEXO 5]

La Técnica de análisis factorial se aplicó para dos DataSet con la finalidad de lograr la reducción de atributos o variables, obteniendo en cada uno los siguientes resultados:

- **Uso de internet en actividades académicas**

Este DataSet estaba compuesto inicialmente de diez atributos, pero con la aplicación del análisis factorial se logró reducir a dos, mismos que son:

“Interacción a través de la Plataforma Virtual” e “Interacción a través de redes sociales”.

- **Uso de internet en actividades de entretenimiento**

Este DataSet estaba compuesto inicialmente de cinco atributos, pero con la aplicación del análisis factorial se logró reducir a tres, mismos que son: “Horas de Chat y redes sociales”, “Utilización de juegos en línea” y “descarga de música y videos”.

Una vez realizado el ejercicio del análisis factorial, y con la finalidad de confirmar la factibilidad de tomar el resultado del análisis factorial como aplicable para la implementación del modelo de Minería de datos en este trabajo, se procede con la utilización del algoritmo de comprobación *DecisionTable*, mismo que nos permite encontrar el coeficiente de correlación entre los atributos que conforman cada DataSet y adicionalmente el porcentaje de error que se tendrá al aplicar la simplificación de atributos o análisis factorial.

El algoritmo de *DecisionTable* para su correcta aplicación solicita el ingreso de los parámetros de: *crossVal* en donde se establece el número de pliegues, en este caso se deja 1 (uno) por default, *displayRules* que es en donde se establece si se necesita imprimir las reglas, dejamos por default “false”, *evaluationMeasure* que es en donde se establece la medida utilizada para evaluar el rendimiento de las combinaciones de atributos utilizados en la tabla de decisión, en este caso se deja la opción de viene por default que es “RMSE (numeric class)”, *search* es donde se ingresa el método de búsqueda que se va a utilizar para encontrar buenas combinaciones de atributos, para este ejercicio se parametrizó el método “BestFirst”. Estos parámetros se pueden ver en él [ANEXO 6].

Como resultado de este ejercicio, se obtuvo que no es factible la aplicación de la técnica de análisis factorial o simplificación de atributos en los DataSet de “uso del internet en actividades académicas” y “uso del internet en actividades de entretenimiento”, esto debido a que el coeficiente de correlación está muy lejos de 1 (uno), en el caso del primer DataSet 0.5761 y en el caso del segundo 0.436. Adicionalmente el error absoluto es demasiado alto, pues para el primer caso se obtuvo 79.11%, y en el segundo caso 86.23%. Estos valores se pueden visualizar en las [Figuras 32 y 33].

```

=== Evaluation on training set ===
=== Summary ===

Correlation coefficient                0.5761
Class complexity | order 0            1797.8583 bits      3.6102 bits/instance
Class complexity | scheme             3050.1675 bits      6.1248 bits/instance
Complexity improvement (Sf)           -1252.3092 bits     -2.5147 bits/instance
Mean absolute error                   5.5389
Root mean squared error               8.5252
Relative absolute error                79.116 %
Root relative squared error           81.7379 %
Total Number of Instances             498

```

**Figura 32.** Resultado aplicación de Tabla de decisión a DataSet "Uso de internet en actividades académicas"

**Fuente:** Generado desde herramienta WEKA

```

=== Evaluation on training set ===
=== Summary ===

Correlation coefficient                0.436
Class complexity | order 0            1975.2267 bits      3.9663 bits/instance
Class complexity | scheme             2382.9954 bits      4.7851 bits/instance
Complexity improvement (Sf)           -407.7687 bits     -0.8188 bits/instance
Mean absolute error                   9.599
Root mean squared error               15.0951
Relative absolute error                86.2329 %
Root relative squared error           89.997 %
Total Number of Instances             498

```

**Figura 33.** Resultado aplicación de Tabla de decisión a DataSet "Uso del internet en actividades de entretenimiento"

**Fuente:** Generado desde herramienta WEKA

#### 6.4.2.2. Clustering (Agrupamiento)

Para realizar el análisis clustering o de agrupamiento, se procedió a utilizar dos algoritmos:

El primero es *EM (Expectation Maximization)* que se caracteriza por iniciar adivinando los parámetros de las distribuciones, los utiliza para calcular las probabilidades de que cada objeto pertenezca a un clúster y usa esas probabilidades para re-estimar los parámetros de las probabilidades, hasta converger. Con lo antes dicho este algoritmo nos permite conocer el número de clústeres más óptimo para el modelo de minería de

datos (Rodríguez et al., 2007), por lo que se ha procedido a aplicar este algoritmo a los DataSet "Uso de internet en actividades académicas" y "Uso de internet en actividades de entretenimiento" obteniendo como resultado que el número de clústeres a aplicar para el primer DataSet es cuatro (4) y para el segunda DataSet es siete (7), esto se puede ver en las [Figuras 34 y 35].

Para aplicar este algoritmo se necesitó realizar las siguientes configuraciones en sus parámetros: si se ubica "true" en *debug*, permite generar información adicional de consola, en este caso la dejaremos por defecto en "false"; *displayModelInOldFormat* se dejara por defecto en "false"; *MaxIterations* que permite indicar el número de iteraciones y se dejara en valor por defecto "100"; *numClusters* que para seleccionar un numero de clúster automáticamente se parametrizo en "-1"; y *seed* que es la semilla del número aleatorio que se utilizará se tomó el valor de "100".

```
=== Model and evaluation on training set ===  
  
Clustered Instances  
  
0      142 ( 29%)  
1      121 ( 24%)  
2      104 ( 21%)  
3      131 ( 26%)  
  
Log likelihood: -30.55058
```

**Figura 34.** Resultado aplicación de algoritmo EM a DataSet "uso del internet en actividades académicas"

**Fuente:** Generado desde herramienta WEKA

```

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

Clustered Instances

0      93 ( 19%)
1     125 ( 25%)
2      39 (  8%)
3      48 ( 10%)
4     118 ( 24%)
5      12 (  2%)
6      63 ( 13%)

Log likelihood: -16.4028

```

**Figura 35.** Resultado aplicación de algoritmo EM a DataSet "uso del internet en actividades de entretenimiento"

**Fuente:** Generado desde herramienta WEKA

El segundo algoritmo es *SimpleKMeans* que está dentro del grupo de técnicas no jerárquicas, y se caracteriza por ser uno de los más veloces y eficientes. También es clasificado como método de particionado y recolocación; y es el algoritmo de agrupamiento más utilizado para proyectos de minería de datos, esto en línea a lo mencionado por (Moore, 2001) y (Jain, 2010). Más detalles respecto a la presente técnica se encuentran en el apartado de marco metodológico [Sección 5.5, punto 5.5.2].

Sin embargo es necesario acotar que uno de los limitantes del algoritmo *SimpleKMeans* es que de antemano hay que proporcionarle el número de clústeres en los que queremos que se segmente la información, tema que lo solventamos con la aplicación del algoritmo *EM*, cuyos resultados nos presentan el número óptimo de clústeres a aplicar en cada uno de los DataSet.

Para realizar la aplicación de este proceso de agrupación mediante el algoritmo *SimpleKMeans*, fue necesario también la aplicación del método *AddCluster* mismo que permite añadir un nuevo atributo nominal que representa el clúster asignado a cada instancia por el algoritmo de agrupamiento especificado, para este caso es *SimpleKMeans*. Los parámetros que se ingresaron respecto al método *AddCluster* fueron *clusterer* que permite el ingreso del algoritmo de agrupamiento a utilizar que para este caso es *SimpleKMeans*, y el parámetro *ignoredAttributeIndices* que permite

ingresar la lista de atributos que se deben ignorar para la aplicación del algoritmo, que para el caso del primer DataSet se dejó el parámetro “first,12-last” y para el segundo DataSet “first-11,17-last”. Ahora bien, dentro de la parametrización del algoritmo *SimpleKMeans* también se realizaron configuraciones importantes como son; *displayStdDevs* que se dejara con el valor por defecto “true” con la finalidad de mostrar la desviación estándar de los atributos, el parámetro de *distanceFunction* permite establecer el método a utilizar para calcular las distancias entre los valores, que para este estudio se utilizara “EuclideanDistance” por la efectividad de la misma, el parámetro *numClusters* establece el número de clústeres resultantes que se requiere obtener, por lo que se ha establecido cuatro para el primer DataSet y siete para el segundo DataSet, el ultimo parámetro *seed* establece la semilla del número aleatorio que se utilizará para el proceso, que en este caso se dejara el valor por defecto “diez” , todas estas configuraciones pueden verse en [ANEXO 7 Y 8].

### Resultados de la aplicación de técnica de agrupación para DataSet “Uso del internet en actividades académicas”

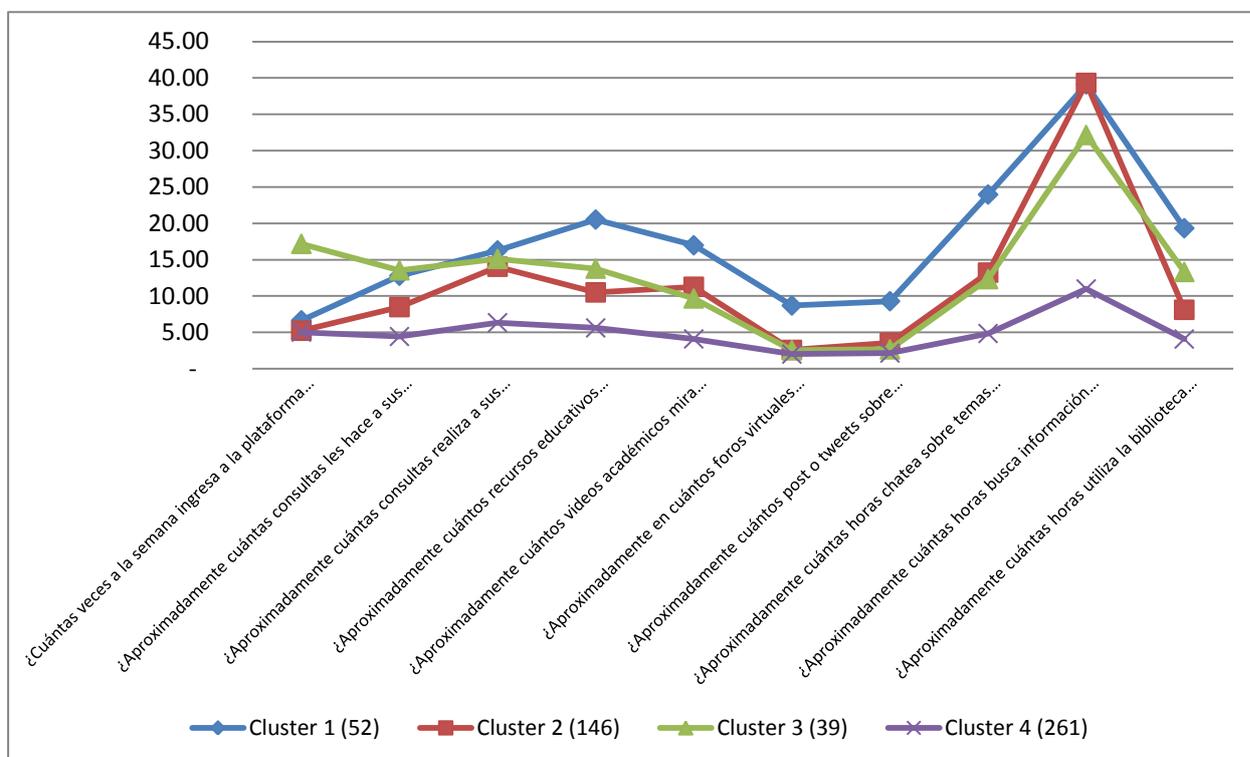
Como resultados obtenidos mediante la aplicación de ésta técnica de agrupación tenemos cuatro grupos, cuyos centroides para cada atributo se pueden observar en la [Tabla 46] y [Figura 36].

**Tabla 46.** Centroides de agrupación de atributos de DataSet Actividades Académicas

Cluster centroids: DataSet (Uso de internet en actividades académicas)	Full Data (498)	Cluster 1 (52)	Cluster 2 (146)	Cluster 3 (39)	Cluster 4 (261)
¿Cuántas veces a la semana ingresa a la plataforma virtual de su universidad?	6.21	6.65	5.23	17.18	5.03
¿Aproximadamente cuántas consultas les hace a sus profesores cada mes?	7.20	12.79	8.48	13.54	4.43
¿Aproximadamente cuántas consultas realiza a sus compañeros por mes?	10.32	16.27	14.03	15.10	6.33
¿Aproximadamente cuántos recursos educativos descarga de la plataforma virtual cada mes?	9.24	20.52	10.50	13.77	5.62
¿Aproximadamente cuántos videos académicos mira en youtube cada mes?	7.98	16.98	11.27	9.67	4.10
¿Aproximadamente en cuántos foros virtuales participa cada mes?	2.92	8.69	2.58	2.56	2.01
¿Aproximadamente cuántos post o tweets sobre temas académicos realiza en las redes sociales por mes?	3.35	9.29	3.58	2.67	2.15
¿Aproximadamente cuántas horas chatea sobre temas académicos cada mes?	9.88	23.96	13.22	12.33	4.84
¿Aproximadamente cuántas horas busca información académica	23.88	39.12	39.32	32.15	10.98

en internet cada mes?					
¿Aproximadamente cuántas horas utiliza la biblioteca virtual de la universidad cada mes?	7.58	19.33	8.11	13.31	4.08

**Fuente:** Elaboración propia a través de Herramienta WEKA



**Figura 36.** Resultado algoritmo *SimpleKmeans* a DataSet "uso del internet en actividades académicas"

**Fuente:** Generado desde herramienta WEKA

El primer clúster está conformado por 52 estudiantes que equivale al 10% de la muestra total, y se caracterizan por estar sobre los demás clústeres o grupos respecto al uso del internet en actividades académicas, lo cual nos indica que este grupo ocupa su tiempo de uso del internet para realizar actividades que aportan a su aprendizaje académico, por lo que les denominaremos como **alto**.

El segundo clúster está conformado por 146 estudiantes que equivale al 29% de la muestra total, y se caracterizan por su interés en la búsqueda de información académica en internet y consultas académicas hacia sus compañeros, adicionalmente interactúan poco con herramientas académicas y redes sociales para su aprendizaje, lo cual nos indica que este grupo utiliza escasamente las herramientas que ofrece el

internet como parte de su proceso de aprendizaje, por lo que les denominaremos como **regular**.

El tercer clúster está conformado por 39 estudiantes que equivale al 8% de la muestra total levantada, y se caracterizan por su alta interacción en la plataforma virtual y su alto uso de la misma para realizar consultas a sus profesores, adicional interactúan poco con herramientas académicas y redes sociales para su aprendizaje, pero lo hacen con mayor frecuencia que el grupo dos, lo que nos lleva a deducir que este grupo interactúa de manera aceptable con las herramientas que ofrece el internet como parte de su proceso de aprendizaje, por lo que les denominaremos como **medio**.

El cuarto clúster está conformado por 261 estudiantes que equivale al 52% de la muestra total levantada, y se caracterizan por su baja interacción con las herramientas que ofrece el internet como parte de su proceso de aprendizaje frente a los demás clústeres o grupos, lo cual nos indica que este grupo casi nunca utiliza el internet para realizar actividades que aportan a su aprendizaje académico, por lo que les denominaremos como **bajo**.

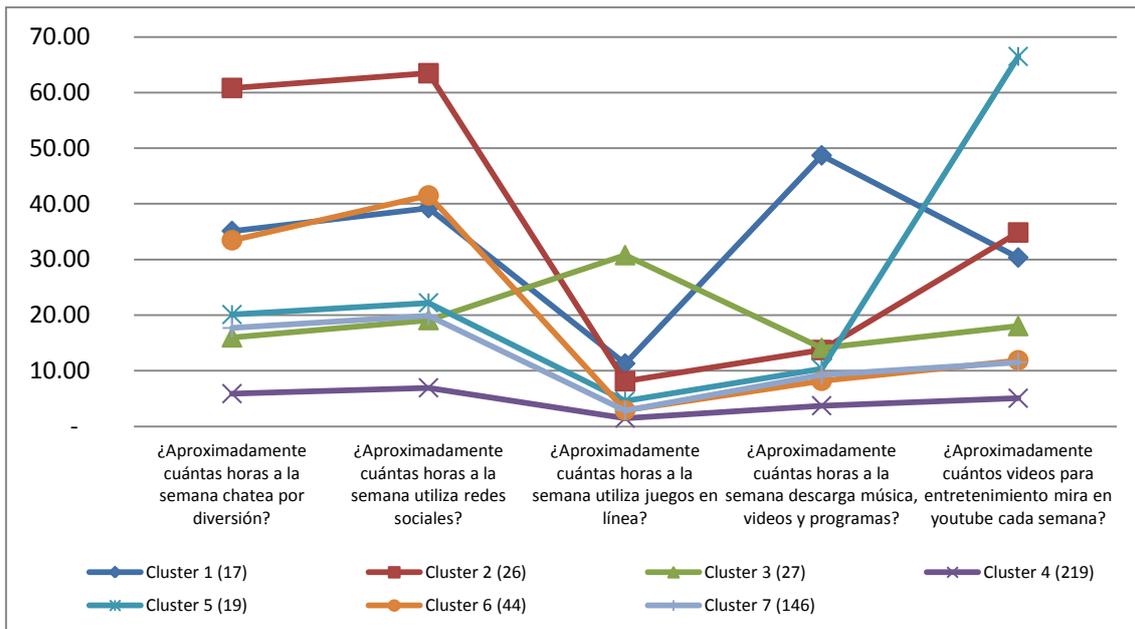
### Resultados de la aplicación de técnica de agrupación para DataSet “Uso del internet en actividades de entretenimiento”

Como resultados obtenidos mediante la aplicación de ésta técnica de agrupación tenemos siete grupos, cuyos centroides para cada atributo se pueden observar en la [Tabla 47] y [Figura 37].

**Tabla 47.** Centroides de agrupación de atributos de DataSet Actividades de entretenimiento

Cluster centroids: DataSet (Uso de internet en actividades de entretenimiento)	F. Data (498)	Cluster 1 (17)	Cluster 2 (26)	Cluster 3 (27)	Cluster 4 (219)	Cluster 5 (19)	Cluster 6 (44)	Cluster 7 (146)
¿Aproximadamente cuántas horas a la semana chatea por diversión?	16.74	35.12	60.81	16.00	5.89	20.11	33.45	17.68
¿Aproximadamente cuántas horas a la semana utiliza redes sociales?	19.08	39.24	63.46	19.11	6.95	22.21	41.50	19.87
¿Aproximadamente cuántas horas a la semana utiliza juegos en línea?	4.41	11.29	8.12	30.78	1.49	4.58	2.89	2.88
¿Aproximadamente cuántas horas a la semana descarga música, videos y programas?	8.62	48.71	13.73	14.11	3.73	10.37	8.20	9.26
¿Aproximadamente cuántos videos para entretenimiento mira en youtube cada semana?	13.03	30.35	34.85	18.04	5.07	66.47	11.89	11.54

**Fuente:** Elaboración propia a través de Herramienta WEKA



**Figura 37.** Resultado algoritmo *SimpleKmeans* a DataSet "uso del internet en actividades de entretenimiento"

**Fuente:** Generado desde herramienta WEKA

El primer clúster está conformado por 17 estudiantes que equivale al 3% de la muestra total, y se caracterizan por estar sobre los demás clústeres o grupos respecto a la descarga de música, videos y programas de computación, y mantienen un nivel medio respecto a juegos en línea y utilización de redes sociales por diversión, lo cual nos indica que este grupo ocupa su tiempo de uso del internet en la descarga de información de entretenimiento, por lo que les denominaremos como grupo de **descargas**.

El segundo clúster está conformado por 26 estudiantes que equivale al 5% de la muestra total, y se caracterizan por estar sobre los demás clústeres o grupos respecto al uso de herramientas de redes sociales y chat por diversión, mantienen un nivel medio respecto a la visualización de videos en internet, y nivel bajo en descargar de información y juegos en línea, lo cual nos indica que este grupo ocupa su tiempo de uso del internet en la web social, por lo que les denominaremos como grupo de **web social**.

El tercer clúster está conformado por 27 estudiantes que equivale al 5% de la muestra total, y se caracterizan por estar sobre los demás clústeres o grupos respecto al uso de herramientas de juegos en línea, y mantienen un nivel bajo respecto al uso de

redes sociales, chat, visualización de videos y descarga de información de entretenimiento, lo cual nos indica que este grupo ocupa su tiempo de uso del internet en los juegos en línea, por lo que les denominaremos como grupo de **jugadores virtuales**.

El cuarto clúster está conformado por 219 estudiantes que equivale al 44% de la muestra total, y se caracterizan por tener el nivel más bajo de uso del internet en actividades de entretenimiento respecto a los demás clústeres o grupos, lo cual nos indica que este grupo considera poco importante el entretenimiento o en su defecto está limitado al acceso de internet por diversos factores, por lo que les denominaremos como grupo **bajo**.

El quinto clúster está conformado por 19 estudiantes que equivale al 4% de la muestra total, y se caracterizan por estar sobre los demás clústeres o grupos respecto al uso de herramientas de visualización de videos de entretenimiento, y mantienen un nivel bajo respecto al uso de redes sociales, chat, juegos en línea y descargar de información de entretenimiento, lo cual nos indica que este grupo ocupa su tiempo de uso del internet en la visualización de videos de entretenimiento, por lo que les denominaremos como grupo de **visualizadores virtuales**.

El sexto clúster está conformado por 44 estudiantes que equivale al 9% de la muestra total, y se caracterizan por tener el nivel medio de uso del internet en actividades de entretenimiento respecto a los demás clústeres o grupos, lo cual nos indica que este grupo emplea su tiempo en de uso del internet de una manera considerable para entretenimiento, por lo que les denominaremos como grupo **medio**.

El séptimo clúster está conformado por 146 estudiantes que equivale al 29% de la muestra total, y se caracterizan por tener el nivel medio-bajo de uso del internet en actividades de entretenimiento respecto a los demás clústeres o grupos, lo cual nos indica que este grupo emplea muy poco tiempo del uso del internet en actividades de entretenimiento, por lo que les denominaremos como grupo **regular**.

#### **6.4.2.3. Regresión logística Multinomial (Predicción de resultados)**

Para aplicar la técnica de regresión logística multinomial se procedió a utilizar el algoritmo *Logistic*, ya que este nos permite aplicar la técnica antes mencionada utilizando un estimador de pliegues. Para aplicar este algoritmo se necesitó realizar las

siguientes configuraciones en sus parámetros: *debug* si se ubica "true", permite generar información adicional de consola, en este caso le dejaremos esta por defecto "false"; *maxIts* permite ingresar el número máximo de iteraciones, en este caso se dejara por defecto en "-1" para que el algoritmo calcule automáticamente las iteraciones que sean necesarias para el ejercicio; y *ridge* que es mediante el cual se establece el valor de pliegues en el log-likelihood, mismo que para esta ejercicio se deberá el parámetro que esta por default.

Mediante esta técnica se aplica las relaciones que nos permiten poder resolver las cuatro hipótesis planteadas para este trabajo de investigación, mismas son:

### **Ingresos económicos del estudiante frente al uso del internet en actividades académica.**

En esta sección se comprueba la hipótesis 1 planteada en este trabajo de investigación que sostiene que el uso del internet en actividades académicas está determinado por el nivel de ingresos familiar que tiene el estudiante. Para este ejercicio se tomara como clase o variable independiente el "ingreso económico del estudiante" y como variable dependiente al resultado del proceso de clusterización respecto al "uso de internet en actividades académicas" que nos generó cuatro categorías: usuarios con uso alto, usuarios con uso medio, usuarios con uso regular y usuarios con uso bajo.

Tomando en cuenta que la variable dependiente está compuesta de cuatro niveles, hace que sea necesaria la utilización de la técnica de regresión logística multinomial, pues esta técnica nos permite trabajar con una variable dependiente que tenga más de dos niveles.

La aplicación el algoritmo *Logistic* nos dio como resultado un valor de índice Kappa de 0.00. Este índice nos muestra el nivel de concordancia entre las categorías pronosticadas por el clasificador y las observadas, en donde: si el valor es negativo quiere decir que la concordancia es muy poca, mientras que si el valor oscila entre cero y uno la concordancia es más alta, hasta llegar a la concordancia perfecta (cuando valor es igual a uno).

Otro indicador que muestra la relación existente entre las variables independiente y dependiente del modelo son los coeficientes que se obtuvo como resultado de la

aplicación de la prueba del cociente de probabilidad (likelihood ratio test), que para todos los casos presenta un valor distinto de cero, al igual que los OR (Odds Ratios) que también presenta un valor distintos de cero, lo que nos indica que tienen una relación significativa, como se lo puede observar en [Figura 38].

```

Logistic Regression with ridge parameter of 1.0E-8
Coefficients...

```

Variable	Class		
	cluster1	cluster2	cluster3
Los ingresos mensuales de su familia son de:=Hasta 1000 d\$lares	0.2985	-0.0245	0.6228
Los ingresos mensuales de su familia son de:=M\$ de 1.500 d\$lares	-0.0956	-0.26	0.0463
Los ingresos mensuales de su familia son de:=Hasta 350 d\$lares	0.2017	-0.0906	-0.7959
Los ingresos mensuales de su familia son de:=Hasta 600 d\$lares	0.0258	0.0711	-0.0555
Los ingresos mensuales de su familia son de:=Hasta 1.500 d\$lares	-0.6346	0.2657	0.2492
Intercept	-1.668	-0.5873	-1.9922

```

Odds Ratios...

```

Variable	Class		
	cluster1	cluster2	cluster3
Los ingresos mensuales de su familia son de:=Hasta 1000 d\$lares	1.3479	0.9758	1.8641
Los ingresos mensuales de su familia son de:=M\$ de 1.500 d\$lares	0.9088	0.771	1.0474
Los ingresos mensuales de su familia son de:=Hasta 350 d\$lares	1.2234	0.9134	0.4512
Los ingresos mensuales de su familia son de:=Hasta 600 d\$lares	1.0261	1.0736	0.9461
Los ingresos mensuales de su familia son de:=Hasta 1.500 d\$lares	0.5302	1.3043	1.2831

**Figura 38.** Coeficientes del modelo de regresión logística multinomial (ingreso económico y uso de Internet en actividades académicas)

**Fuente:** Generado desde herramienta WEKA

Los resultados mostrados en la [Figura 38] nos dejan apreciar que el OR (Odds Ratios), al ser un estudiante parte del grupo **regular** (cluster2) con relación al grupo **bajo** (cluster4) respecto al uso de internet para actividades académicas, disminuye 1.0948 veces cuando un estudiante tiene un ingreso mensual hasta de 350,00 dólares, en comparación con los estudiantes que pertenecen a los otros rangos de ingresos (1 / (OR = 0.9134)); aumenta 1.0736 veces cuando un estudiante tiene un ingreso mensual hasta de 600,00 dólares, en comparación con los estudiantes que pertenecen a los otros rangos de ingresos (OR = 1.0736); disminuye 1.0248 veces cuando un estudiante tiene un ingreso mensual hasta de 1000,00 dólares, en comparación con los estudiantes que pertenecen a los otros rangos de ingresos (1 / (OR = 0.9758)); aumenta 1.3043 veces cuando un estudiante tiene un ingreso mensual hasta de 1500,00 dólares, en comparación con los estudiantes que pertenecen a los otros rangos de ingresos (OR1 = 1.3043); disminuye 1.2970 veces cuando un estudiante tiene un ingreso mensual mayor a 1500,00 dólares, en comparación con los estudiantes que pertenecen a los otros rangos de ingresos (1 / (OR = 0.771)).

Adicional en la [Figura 39] se puede evidenciar también como resultado los indicadores de: verdaderos positivos, falsos positivos, medida de precisión, recall, medida-F, y también se puede visualizar que todos los registros están clasificados dentro del clúster cuatro “nivel bajo de uso internet en actividades académicas”.

```

=== Detailed Accuracy By Class ===

      TP Rate   FP Rate   Precision   Recall   F-Measure   ROC Area   Class
      0         0         0           0         0           0.484     cluster1
      0         0         0           0         0           0.476     cluster2
      0         0         0           0         0           0.553     cluster3
      1         1         0.524       1         0.688       0.465     cluster4
Weighted Avg.  0.524     0.524       0.275     0.524     0.36        0.477
    
```

**Figura 39.** Resultado aplicación de técnica de regresión logística multinomial (ingreso económico y uso de Internet en actividades académicas)

**Fuente:** Generado desde herramienta WEKA

También este modelo a logrado el 52.41% de instancias correctamente clasificadas del universo de la muestra, es decir 261 registros de 498, mismas que están todas clasificadas dentro del clúster cuatro, como lo muestra la [Tabla 48].

**Tabla 48.** Matriz de confusión ingreso económico y uso del internet en actividades académicas

Variables Depend.	Predicción				% Correcto
	Uso Alto	Uso Regular	Uso Medio	Uso Bajo	
Uso Alto	0	0	0	52	0.00%
Uso Regular	0	0	0	146	0.00%
Uso Medio	0	0	0	39	0.00%
Uso Bajo	0	0	0	261	100.00%

**Fuente:** Elaboración propia a través de Herramienta WEKA

### **Ingresos económicos del estudiante frente al uso del internet en actividades de entretenimiento.**

En esta sección se comprueba la hipótesis 2 planteada en este trabajo de investigación que sostiene que el uso del internet en actividades de entretenimiento está determinado por el nivel de ingresos familiar que tiene el estudiante. Para este ejercicio se tomara como clase o variable independiente el “ingreso económico del estudiante” y como variable dependiente al resultado del proceso de clusterización respecto al “uso de internet en actividades de entretenimiento”, mismo que nos generó

siete categorías: usuarios con uso en descargas, usuarios con uso en web social, usuarios con uso en juegos virtuales, usuarios con uso bajo, usuario con uso en visualización de videos, usuarios con uso medio y usuarios con uso regular. Cabe mencionar que dentro de los grupos de clusterización no existe el de uso alto del internet en actividades de entretenimiento, toda vez que los usuarios que hacen uso alto del internet, están agrupados como especializados.

Tomando en cuenta que la variable dependiente está compuesta de siete niveles, hace que sea necesaria la utilización de la técnica de regresión logística multinomial, pues esta técnica nos permite trabajar con una variable dependiente que tenga más de dos niveles.

La aplicación el algoritmo *Logistic* nos dio como resultado un valor de índice Kappa de 0.00. Este índice nos muestra la concordancia entre las categorías pronosticadas por el clasificador y las observadas, en donde: si el valor es negativo quiere decir que la concordancia es muy poca, mientras que si el valor oscila entre cero y uno la concordancia es más alta, hasta llegar a la concordancia perfecta (cuando valor es igual a uno).

Otro indicador que muestra la relación existente entre las variables independiente y dependiente son los coeficientes del modelo, que se obtuvo como resultado de la aplicación de la prueba del cociente de probabilidad (likelihood ratio test), que para todos los casos presenta un valor distinto de cero, al igual que los OR (Odds Ratios) que también presenta un valor distintos de cero, lo que nos indica que tienen una relación significativa, como se lo puede observar en [Figura 40].

Logistic Regression with ridge parameter of 1.0E-8  
Coefficients...

Variable	Class					
	cluster1	cluster2	cluster3	cluster4	cluster5	cluster6
Los ingresos mensuales de su familia son de:=Hasta 1000 dÃlares	3.6312	0.4684	0.729	0.0188	0.49	0.0806
Los ingresos mensuales de su familia son de:=MÃs de 1.500 dÃlares	3.1333	0.5096	-0.375	-0.0525	0.3977	0.4582
Los ingresos mensuales de su familia son de:=Hasta 350 dÃlares	1.3416	-1.5699	-0.375	0.281	-0.2955	-0.3685
Los ingresos mensuales de su familia son de:=Hasta 600 dÃlares	2.5407	0.3224	-0.1568	-0.2608	-0.8881	-0.1037
Los ingresos mensuales de su familia son de:=Hasta 1.500 dÃlares	-12.6663	0.6128	0.0159	-0.0065	0.6062	0.1559
Intercept	-4.8679	-1.9565	-1.7651	0.3974	-2.1323	-1.212

Odds Ratios...

Variable	Class					
	cluster1	cluster2	cluster3	cluster4	cluster5	cluster6
Los ingresos mensuales de su familia son de:=Hasta 1000 dÃlares	37.7574	1.5975	2.0729	1.019	1.6324	1.0839
Los ingresos mensuales de su familia son de:=MÃs de 1.500 dÃlares	22.9506	1.6646	0.6873	0.9488	1.4884	1.5812
Los ingresos mensuales de su familia son de:=Hasta 350 dÃlares	3.8251	0.2081	0.6873	1.3244	0.7442	0.6918
Los ingresos mensuales de su familia son de:=Hasta 600 dÃlares	12.6881	1.3804	0.8549	0.7704	0.4114	0.9015
Los ingresos mensuales de su familia son de:=Hasta 1.500 dÃlares	0	1.8455	1.016	0.9935	1.8335	1.1687

**Figura 40.** Coeficientes del modelo de regresión logística multinomial (ingreso económico y uso de Internet en actividades de entretenimiento)

**Fuente:** Generado desde herramienta WEKA

Los resultados mostrados en la [Figura 40] nos dejan apreciar que el OR (Odds Ratios), al ser un estudiante parte del grupo **bajo** (cluster4) con relación al grupo **regular** (cluster7) respecto al uso de internet para actividades de entretenimiento, aumenta 1.3244 veces cuando un estudiante tiene un ingreso mensual hasta de 350,00 dólares, en comparación con los estudiantes que pertenecen a los otros rangos de ingresos (OR = 1.3244); disminuye 1.2980 veces cuando un estudiante tiene un ingreso mensual hasta de 600,00 dólares, en comparación con los estudiantes que pertenecen a los otros rangos de ingresos ( $1 / (OR = 0.7704)$ ); aumenta 1.0190 veces cuando un estudiante tiene un ingreso mensual hasta de 1000,00 dólares, en comparación con los estudiantes que pertenecen a los otros rangos de ingresos (OR = 1.0190); disminuye 1.0065 veces cuando un estudiante tiene un ingreso mensual hasta de 1500,00 dólares, en comparación con los estudiantes que pertenecen a los otros rangos de ingresos ( $1 / (OR = 0.9935)$ ); disminuye 1.0540 veces cuando un estudiante tiene un ingreso mensual mayor a 1500,00 dólares, en comparación con los estudiantes que pertenecen a los otros rangos de ingresos ( $1 / (OR = 0.9488)$ ).

Adicional en la [Figura 41] se puede evidenciar también como resultado los indicadores de: verdaderos positivos, falsos positivos, medida de precisión, recall, medida-F, y también se puede visualizar que todos los registros están clasificados dentro del clúster cuatro “nivel bajo de uso internet en actividades de entretenimiento”.

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0	0	0	0	0	0.653	cluster1
	0	0	0	0	0	0.442	cluster2
	0	0	0	0	0	0.503	cluster3
	1	1	0.44	1	0.611	0.483	cluster4
	0	0	0	0	0	0.48	cluster5
	0	0	0	0	0	0.45	cluster6
	0	0	0	0	0	0.447	cluster7
Weighted Avg.	0.44	0.44	0.193	0.44	0.269	0.474	

**Figura 41.** Resultado aplicación de técnica de regresión logística multinomial (ingreso económico y uso de Internet en actividades de entretenimiento)

**Fuente:** Generado desde herramienta WEKA

También este modelo a logrado el 43.98% de instancias correctamente clasificadas del universo de la muestra, es decir 219 registros de 498, mismas que están todas clasificadas dentro del clúster cuatro, como lo muestra la [Tabla 49].

**Tabla 49. Matriz** de confusión (ingreso económico y uso del internet en actividades de entretenimiento)

Variables Dependientes	Predicción							% Correcto
	Uso en Descargas	Uso en Web Social	Uso en Juegos Virtuales	Uso Bajo en Todo	Uso en Ver videos	Uso Medio en Todo	Uso Regular en Todo	
Uso en Descargas	0	0	0	17	0	0	0	0.00%
Uso en Web Social	0	0	0	26	0	0	0	0.00%
Uso en Juegos Virt.	0	0	0	27	0	0	0	0.00%
Uso Bajo en Todo	0	0	0	219	0	0	0	100.00%
Uso en Ver videos	0	0	0	19	0	0	0	0.00%
Uso Medio en Todo	0	0	0	44	0	0	0	0.00%
Uso Regular en Todo	0	0	0	146	0	0	0	0.00%

**Fuente:** Elaboración propia a través de Herramienta WEKA

### Uso del internet en actividades académica frente al rendimiento académico.

En esta sección se comprueba la hipótesis 3 planteada en este trabajo de investigación que sostiene que el rendimiento académico de los estudiantes de nivel superior está determinado por el uso del internet en actividades académicas. Para este ejercicio se tomara como clase o variable independiente al resultado del proceso de clusterización respecto al “uso de internet en actividades académicas” y como variable dependiente el “rendimiento académico de los estudiantes”, mismo que se agrupo en cuatro categorías: estudiantes sin ninguna materia reprobada en categoría uno, estudiantes con una materia reprobada en categoría dos, estudiantes con dos materias reprobadas en categoría tres, y estudiantes con tres o más materias reprobadas en categoría cuatro.

Tomando en cuenta que la variable dependiente está compuesta de cuatro niveles, hace que sea necesaria la utilización de la técnica de regresión logística multinomial, pues esta técnica nos permite trabajar con una variable dependiente que tenga más de dos niveles.

La aplicación del algoritmo *Logistic* nos dio como resultado un valor de índice Kappa de 0.00. Este índice nos muestra la concordancia entre las categorías pronosticadas por el clasificador y las observadas, en donde: si el valor es negativo quiere decir que la concordancia es muy poca, mientras que si el valor oscila entre cero y uno la concordancia es más alta, hasta llegar a la concordancia perfecta (cuando el valor es igual a uno).

Otro indicador que muestra la relación existente entre las variables independiente y dependiente son los coeficientes del modelo, que se obtuvo como resultado de la aplicación de la prueba del cociente de probabilidad (likelihood ratio test), que para todos los casos presenta un valor distinto de cero, al igual que los OR (Odds Ratios) que en la mayoría de valores presenta un valor distinto de cero, lo que nos indica que tienen una relación significativa, como se lo puede observar en [Figura 42].

```

Logistic Regression with ridge parameter of 1.0E-8
Coefficients...

```

Variable	Class		
	1	2	3
clusterig_uso_internet_actividades_academicas=cluster1	-0.2083	0.0473	-14.8067
clusterig_uso_internet_actividades_academicas=cluster2	0.1371	0.2296	2.4271
clusterig_uso_internet_actividades_academicas=cluster3	0.1633	0.5173	2.204
clusterig_uso_internet_actividades_academicas=cluster4	-0.0831	-0.3582	2.8971
Intercept	3.204	1.5622	-2.204

```

Odds Ratios...

```

Variable	Class		
	1	2	3
clusterig_uso_internet_actividades_academicas=cluster1	0.812	1.0484	0
clusterig_uso_internet_actividades_academicas=cluster2	1.147	1.2581	11.3263
clusterig_uso_internet_actividades_academicas=cluster3	1.1774	1.6775	9.0611
clusterig_uso_internet_actividades_academicas=cluster4	0.9203	0.6989	18.1221

**Figura 42.** Coeficientes del modelo de regresión logística multinomial (uso de Internet en actividades académicas y rendimiento académico)

**Fuente:** Generado desde herramienta WEKA

Los resultados mostrados en la [Figura 42] nos dejan apreciar que el OR (Odds Ratios), al ser un estudiante parte del grupo que **no reprueba ninguna materia** con relación al grupo **reprueba tres o más materias**, disminuye 1.2315 veces cuando un estudiante tiene un **alto** nivel uso de internet en actividades académicas, en

comparación con los estudiantes que pertenecen a los otros niveles de uso (1 / (OR = 0.812)); aumenta 1.1470 veces cuando un estudiante tiene un nivel **regular** de uso de internet en actividades académicas, en comparación con los estudiantes que pertenecen a los otros niveles de uso (OR = 1.1470); aumenta 1.1774 veces cuando un estudiante tiene un nivel **medio** de uso de internet en actividades académicas, en comparación con los estudiantes que pertenecen a los otros niveles de uso (OR = 1.1774); disminuye 1.0866 veces cuando un estudiante tiene un nivel **bajo** de uso de internet en actividades académicas, en comparación con los estudiantes que pertenecen a los otros niveles de uso (1 / (OR = 0.9203)).

Adicional en la [Figura 43] se puede evidenciar también como resultado los indicadores de: verdaderos positivos, falsos positivos, medida de precisión, recall, medida-F, y también se puede visualizar que todos los registros están clasificados dentro de la categoría uno “Estudiantes sin ninguna materia reprobada”.

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
1	1	1	0.775	1	0.873	0.355	1
0	0	0	0	0	0	0.497	2
0	0	0	0	0	0	0.557	3
0	0	0	0	0	0	0.296	4
Weighted Avg.	0.775	0.775	0.601	0.775	0.677	0.384	

**Figura 43.** Resultado aplicación de técnica de regresión logística multinomial (uso de Internet en actividades académicas y rendimiento académico)

**Fuente:** Generado desde herramienta WEKA

También este modelo a logrado el 77.51% de instancias correctamente clasificadas del universo de la muestra, es decir 386 registros de 498, mismas que están todas clasificadas dentro del clúster cuatro, como lo muestra la [Tabla 50].

**Tabla 50.** Matriz de confusión (uso del internet en actividades académicas y rendimiento académico)

Variables Dependientes	Predicción				% Correcto
	Cero materias reprobadas	Una materia reprobadas	Dos materias reprobadas	Tres o más materias reprobadas	
Cero materias reprobadas	386	0	0	0	100.00%
Una materia reprobadas	72	0	0	0	0.00%
Dos materias reprobadas	24	0	0	0	0.00%
Tres o más materias	16	0	0	0	0.00%

reprobadas					
------------	--	--	--	--	--

**Fuente:** Elaboración propia a través de Herramienta WEKA

### **Uso del internet en actividades de entretenimiento frente al rendimiento académico.**

En esta sección se comprueba la hipótesis 4 planteada en este trabajo de investigación que sostiene que el rendimiento académico de los estudiantes de nivel superior está determinado por el uso del internet en actividades de entretenimiento. Para este ejercicio se tomara como variable independiente al resultado del proceso de clusterización respecto al “uso de internet en actividades de entretenimiento” y como variable dependiente el “rendimiento académico de los estudiantes”, mismo que se agrupo en cuatro categorías: estudiantes sin ninguna materia reprobada en categoría uno, estudiantes con una materia reprobada en categoría dos, estudiantes con dos materias reprobadas en categoría tres, y estudiantes con tres o más materias reprobadas en categoría cuatro.

Tomando en cuenta que la variable dependiente está compuesta de cuatro niveles, hace que sea necesaria la utilización de la técnica de regresión logística multinomial, pues esta técnica nos permite trabajar con una variable dependiente que tenga más de dos niveles.

La aplicación el algoritmo *Logistic* nos dio como resultado un valor de índice Kappa de 0.00. Este índice nos muestra la concordancia entre las categorías pronosticadas por el clasificador y las observadas, en donde: si el valor es negativo quiere decir que la concordancia es muy poca, mientras que si el valor oscila entre cero y uno la concordancia es más alta, hasta llegar a la concordancia perfecta (cuando valor es igual a uno).

Otro indicador que muestra la relación existente entre las variables independiente y dependiente son los coeficientes del modelo, que se obtuvo como resultado de la aplicación de la prueba del cociente de probabilidad (likelihood ratio test), que para todos los casos presenta un valor distinto de cero, al igual que los OR (Odds Ratios) que en su mayoría presentan un valor distintos de cero, lo que nos indica que tienen una relación significativa, como se lo puede observar en [Figura 44].

Logistic Regression with ridge parameter of 1.0E-8  
Coefficients...

Variable	Class		
	1	2	3
clustering_uso_internet_actividades_entretenimiento=cluster1	17.1087	16.8032	4.1439
clustering_uso_internet_actividades_entretenimiento=cluster2	17.3419	17.8473	4.56
clustering_uso_internet_actividades_entretenimiento=cluster3	15.0587	14.7312	18.4921
clustering_uso_internet_actividades_entretenimiento=cluster4	-3.3587	-3.3254	0.4664
clustering_uso_internet_actividades_entretenimiento=cluster5	-5.0192	-4.7762	-1.0152
clustering_uso_internet_actividades_entretenimiento=cluster6	-3.9465	-4.4885	-17.5683
clustering_uso_internet_actividades_entretenimiento=cluster7	-4.1734	-4.0365	0.1299
Intercept	6.891	5.1816	0.3221

Odds Ratios...

Variable	Class		
	1	2	3
clustering_uso_internet_actividades_entretenimiento=cluster1	26929488.9536	19838869.6797	63.0474
clustering_uso_internet_actividades_entretenimiento=cluster2	34001470.1326	56359731.7284	95.5867
clustering_uso_internet_actividades_entretenimiento=cluster3	3466745.5716	2498420.1582	107407653.0137
clustering_uso_internet_actividades_entretenimiento=cluster4	0.0348	0.036	1.5942
clustering_uso_internet_actividades_entretenimiento=cluster5	0.0066	0.0084	0.3623
clustering_uso_internet_actividades_entretenimiento=cluster6	0.0193	0.0112	0
clustering_uso_internet_actividades_entretenimiento=cluster7	0.0154	0.0177	1.1387

**Figura 44.** Coeficientes del modelo de regresión logística multinomial (uso de Internet en actividades de entretenimiento y rendimiento académico)

**Fuente:** Generado desde herramienta WEKA

Los resultados mostrados en la [Figura 44] nos dejan apreciar que el OR (Odds Ratios), al ser un estudiante parte del grupo que **no reprueba ninguna materia** con relación al grupo **reprueba tres o más materia**, aumenta 26929488.9536 veces cuando un estudiante tiene un nivel de uso de internet especializado en **descargas**, en comparación con los estudiantes que pertenecen a los otros grupos de uso (OR = 26929488.9536); aumenta 34001470.1326 veces cuando un estudiante tiene un nivel de uso de internet especializado en **web social**, en comparación con los estudiantes que pertenecen a los otros grupos de uso (OR = 34001470.1326); aumenta 3466745.5716 veces cuando un estudiante tiene un nivel de uso de internet especializado en **juegos virtuales**, en comparación con los estudiantes que pertenecen a los otros grupos de uso (OR = 3466745.5716); disminuye 28.74 veces cuando un estudiante tiene un nivel **bajo** de uso de internet en actividades de entretenimiento, en comparación con los estudiantes que pertenecen a los otros niveles o grupos de uso ( $1 / (OR = 0.0348)$ ); disminuye 151.52 veces cuando un estudiante tiene un nivel de uso de internet especializado en **ver videos**, en comparación con los estudiantes que pertenecen a los otros niveles o grupos de uso ( $1 / (OR = 0.0066)$ ); disminuye 51.81 veces cuando un estudiante tiene un nivel **medio** de uso de internet en actividades de entretenimiento, en comparación con los estudiantes que pertenecen a los otros niveles o grupos de uso ( $1 / (OR = 0.0193)$ ); disminuye 64.94 veces cuando un estudiante tiene un nivel **regular** de uso de internet

en actividades de entretenimiento, en comparación con los estudiantes que pertenecen a los otros niveles o grupos de uso ( $1 / (OR = 0.0154)$ ).

Adicional en la [Figura 45] se puede evidenciar también como resultado los indicadores de: verdaderos positivos, falsos positivos, medida de precisión, recall, medida-F, y también se puede visualizar que todos los registros están clasificados dentro de la categoría uno “Estudiantes sin ninguna materia reprobada”.

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	1	1	0.775	1	0.873	0.513	1
	0	0	0	0	0	0.439	2
	0	0	0	0	0	0.526	3
	0	0	0	0	0	0.538	4
Weighted Avg.	0.775	0.775	0.601	0.775	0.677	0.504	

**Figura 45.** Resultado aplicación de técnica de regresión logística multinomial (uso de Internet en actividades de entretenimiento y rendimiento académico)

**Fuente:** Generado desde herramienta WEKA

También este modelo a logrado el 77.51% de instancias correctamente clasificadas del universo de la muestra, es decir 386 registros de 498, mismas que están todas clasificadas dentro del clúster cuatro, como lo muestra la [Tabla 51].

**Tabla 51.** Matriz de confusión (uso del internet en actividades de entretenimiento y rendimiento académico)

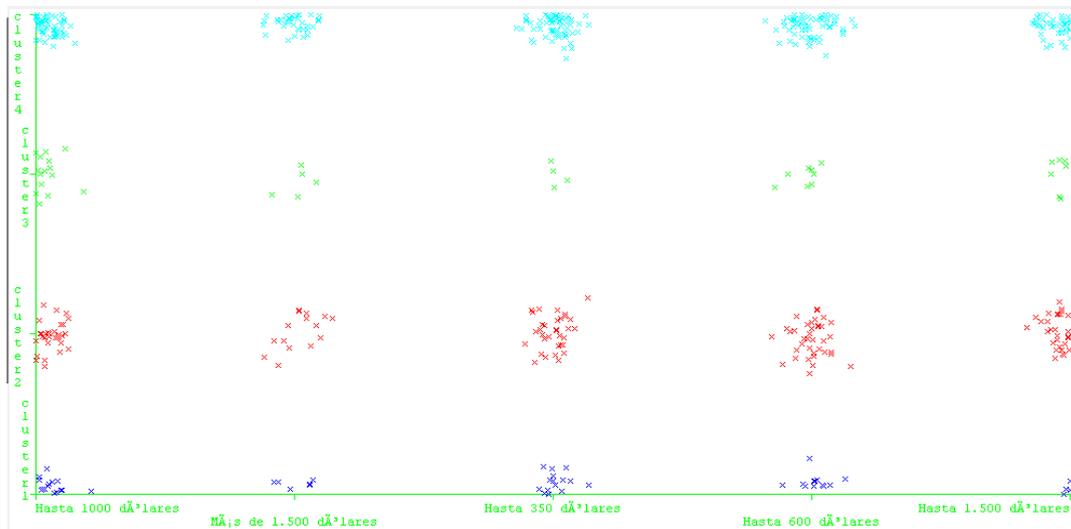
Variables Dependientes	Predicción				% Correcto
	Cero materias reprobadas	Una materia reprobadas	Dos materias reprobadas	Tres o más materias reprobadas	
Cero materias reprobadas	386	0	0	0	100.00%
Una materia reprobadas	72	0	0	0	0.00%
Dos materias reprobadas	24	0	0	0	0.00%
Tres o más materias reprobadas	16	0	0	0	0.00%

**Fuente:** Elaboración propia a través de Herramienta WEKA

#### 6.4.2.4. Gráficos de dispersión

**Distribución de relación entre ingreso económico - uso de Internet en actividades académicas**

En la [Figura 46] se observa que la mayoría de los registros están clasificados dentro del clúster cuatro “Nivel bajo de uso internet en actividades académicas”.

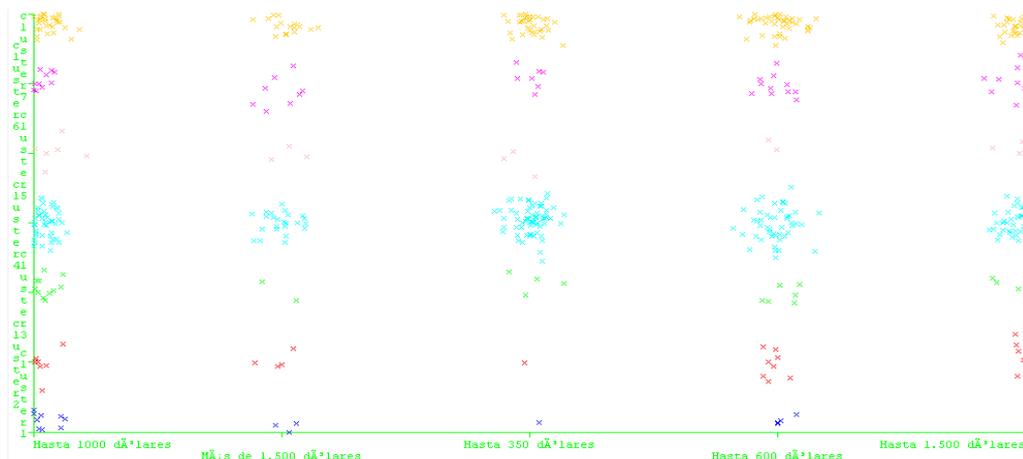


**Figura 46.** Grafica de relación de ingreso económico - uso de Internet en actividades académicas

**Fuente:** Generado desde herramienta WEKA

### Distribución de relación entre ingreso económico - uso de Internet en actividades de entretenimiento

En la [Figura 47] se observa que la mayoría de los registros están clasificados dentro del clúster cuatro “Nivel bajo de uso internet en actividades de entretenimiento”.

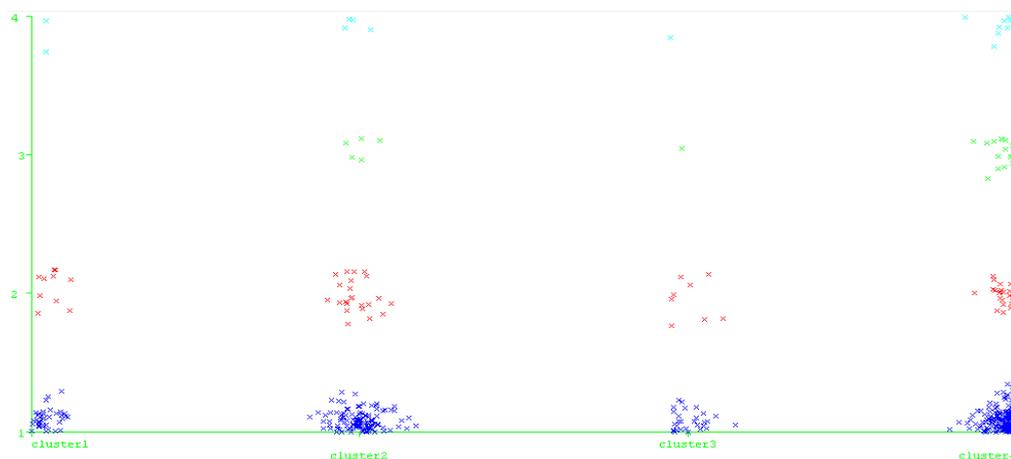


**Figura 47.** Grafica de relación de ingreso económico - uso de Internet en actividades de entretenimiento

**Fuente:** Generado desde herramienta WEKA

## Distribución de relación entre el uso de Internet en actividades académicas – rendimiento académico

En la [Figura 48] se observa que la mayoría de los registros están clasificados dentro de la categoría uno “Estudiantes sin ninguna materia reprobada”.

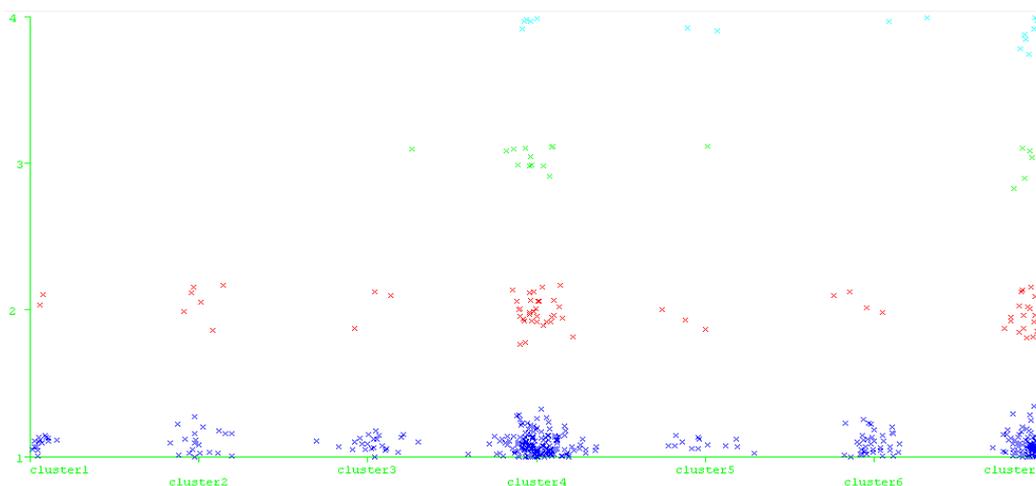


**Figura 48.** Grafica de relación del uso de Internet en actividades académicas - rendimiento académico

**Fuente:** Generado desde herramienta WEKA

## Distribución de relación entre el uso de Internet en actividades de entretenimiento – rendimiento académico

En la [Figura 49] se observa que la mayoría de los registros están clasificados dentro de la categoría uno “Estudiantes sin ninguna materia reprobada”.



**Figura 49.** Grafica de relación del uso de Internet en actividades de entretenimiento - rendimiento académico

**Fuente:** Generado desde herramienta WEKA

### 6.5. Fase V – Evaluación

En esta fase el modelo obtenido es evaluado, tomando en consideración los objetivos del la minería de datos (objetivos del negocio), mismos que fueron detallados en la [Sección 6.1, punto 6.1.1].

El modelo implementado a través de la utilización de las diferentes técnicas de minerías de datos y los algoritmos que nos permitieron aplicarlas, mismos que fueron detallados en la [Sección 6.4, punto 6.4.1], nos han ayudado a evidenciar que las hipótesis planteadas no se cumplen para los estudiantes de la Pontificia Universidad Católica del Ecuador – Seda Ibarra. Mediante la técnica de *Análisis factorial* se pretendía reducir el número de atributos o variables a utilizar, sin embargo al aplicar el método de comprobación de la *tabla de decisión* se evidencio que no puede ser aplicable debido a que su coeficiente de correlación estaba muy alejado de uno (1). También se procedió con la aplicación de la técnica *clustering* a través de dos algoritmos: el primero que nos ayudó a poder saber el número de grupos óptimos para el modelo, y el segundo que nos ayudó a la aplicación de la clusterización respecto a

los DataSet “uso de internet en actividades académicas” y “uso de internet en actividades de entretenimiento”. Finalmente se aplicó la técnica de *regresión logística*, misma que permitió relacionar las variables resultantes de las técnicas aplicadas anteriormente, alineados a la demostración de las hipótesis planteadas para este trabajo de investigación.

Adicionalmente para la implementación del modelo fue necesaria la utilización de métodos y filtros dentro de la herramienta WEKA, en su mayor parte para poder categorizar el atributo de “rendimiento académico”.

## **CAPÍTULO VII - DISCUSIÓN Y CONCLUSIONES**

## 7. Discusión y Conclusiones

En este apartado se discuten los resultados obtenidos relacionados a las hipótesis planteadas para esta investigación, comenzando por la incidencia que tiene el nivel de ingreso frente al uso del internet tanto para actividades académicas, como para actividades de entretenimiento, y luego el efecto de esta incidencia frente al rendimiento académico de los estudiantes.

También de acuerdo a los resultados obtenidos se instituyen conclusiones orientadas a las hipótesis planteadas y a modelos de Minería de Datos creados en este trabajo.

### 7.1. Incidencia del nivel de Ingreso frente al uso de internet en actividades académicas.

La discusión que se presenta en esta sección está alineada a los resultados obtenidos relacionados con la hipótesis 1, misma que sostiene que el nivel de ingreso incide en el uso de internet en actividades académicas.

Para lograr la determinación del uso de internet en actividades académicas se intentó aplicar la técnica de análisis factorial, sin embargo no fue factible debido a que en la comprobación de aplicación de esta técnica, el coeficiente de correlación está muy lejos de uno (1). Luego se aplicó la técnica de clustering tendiendo como resultado cuatro grupos académicos (alto, medio, regular y bajo), y cada uno de ellos se caracteriza por la valoración que presentó respecto a las diferentes variables del DataSet de “Uso del internet en actividades académicas”.

El grupo académico **alto**, está conformado por el 10% de la muestra total, y tiene los valores más altos en todas las variables, especialmente en las variables relacionadas con redes sociales y plataforma virtual, esta igualdad entre las variables nos indica que este grupo ocupa su tiempo de uso del internet en actividades académicas de una manera equilibrada.

El grupo académico **regular**, está conformado por el 29% de la muestra total, y si bien no presenta los valores más bajos de entre los grupos, los mismos no son significativos, salvo el caso de *búsqueda de información en internet* en la cual presenta el valor más alto. Esta característica de especialización en la *búsqueda de información en internet* que tiene este grupo, y el bajo interés por las herramientas que ofrece

actualmente la tecnología, nos indica que este grupo ocupa su tiempo de uso del internet en actividades académicas de una manera desequilibrada y bastante baja.

El grupo académico **medio**, está conformado por el 8% de la muestra total, y presenta los valores medios en la mayoría de variables, salvo el caso de las variables *ingreso en la plataforma virtual* y *consultas a sus profesores* en las cuales presenta los valores más altos. Esta característica de especialización en el *ingreso en la plataforma virtual* y en *consultas a sus profesores* que tiene este grupo, y el nivel medio en los valores de las demás variables, nos indica que este grupo ocupa su tiempo de uso del internet en actividades académicas de una manera desequilibrada.

El grupo académico **bajo**, está conformado por el 52% de la muestra total, y tiene los valores más bajos en todas las variables, lo cual nos indica que el tiempo de uso de internet que ocupa este grupo en actividades académicas es equilibrado pero demasiado bajo.

Con la finalidad de poder obtener la incidencia que tiene el ingreso económico con el uso de internet en actividades académicas, se procedió a aplicar la técnica de regresión logística multinomial, obteniendo como resultado que para todos los rangos de ingresos que tienen los estudiantes de la Pontificia Universidad Católica del Ecuador – Sede Ibarra, indican en un uso **bajo** de internet en actividades académicas. Esto nos indica que la hipótesis planteada: “el nivel de ingresos incide sobre el uso de internet en actividades académicas” para los estudiantes de la Pontificia Universidad Católica del Ecuador – Sede Ibarra es **nula**, toda vez que independientemente del nivel de ingresos que tengan los estudiantes, los mismos presentan un nivel bajo de uso de internet en actividades académicas.

En estudio emprendido por (Torres, 2012) ha obtenido resultados significativos sobre esta hipótesis, aplicada en otras universidades de Ecuador, y el estudio internacional de (Rojas et al., 2015) nos muestra que existe una desventaja en los hogares con menos ingresos económicos por tener menor posibilidad de acceso a internet, lo cual incide en el uso del internet, sin embargo en este trabajo de investigación se presenta a esta hipótesis como nula, ya que el 52% de la muestra total de estudiantes presenta un nivel bajo, y el 29% presenta un nivel regular de uso de internet en actividades académicas, que sumando tenemos que el 81% de estudiantes no tienen un uso significativo de internet, lo que hace que todos niveles de ingreso económico de los estudiantes apunten hacia el bajo uso de internet en actividades académicas, y esto

nos hace concluir que el ingreso económico de los estudiantes de la Pontificia Universidad Católica del Ecuador – Sede Ibarra, no incide en el uso de internet en actividades académicas por parte de los mismos.

## **7.2. Incidencia del nivel de Ingreso frente al uso de internet en actividades de entretenimiento.**

La discusión que se presenta en esta sección está alineada a los resultados obtenidos relacionados con la hipótesis 2, misma que sostiene que el nivel de ingreso incide en el uso de internet en actividades de entretenimiento.

Para lograr la determinación del uso de internet en actividades de entretenimiento se intentó aplicar la técnica de análisis factorial, sin embargo no fue factible debido a que en la comprobación de aplicación de esta técnica, el coeficiente de correlación está muy lejos de uno (1). Luego se aplicó la técnica de Clustering tendiendo como resultado siete grupos de entretenimiento (descargas, web social, juegos virtuales, bajo, visualizador de videos, medio y regular), y cada uno de ellos se caracteriza por la valoración que presentó respecto a las diferentes variables del DataSet de “Uso del internet en actividades de entretenimiento”.

Los grupos de entretenimiento **descargas, web social, juegos virtuales y visualizador de videos**, está conformado por el 17% de la muestra total, y tienen los valores más altos en cada variable en la que se especializan respectivamente. Esta característica de especialización en una actividad de uso de internet para entretenimiento y el bajo uso de las demás herramientas, nos indica que este grupo ocupa su tiempo de uso del internet para una actividad de entretenimiento específico.

El grupo de entretenimiento **regular**, está conformado por el 29% de la muestra total, y si bien no presenta los valores más bajos de entre los grupos, los mismos no son significativos, lo cual nos indica que este grupo ocupa su tiempo de uso de internet en todas las actividades de entretenimiento con un nivel regular (medio-bajo).

El grupo de entretenimiento **medio**, está conformado por el 9% de la muestra total, y si bien no presenta los valores más altos de entre los grupos, los mismos son significativos, lo cual nos indica que este grupo ocupa su tiempo de uso de internet en todas las actividades de entretenimiento con un nivel medio.

El grupo de entretenimiento **bajo**, está conformado por el 44% de la muestra total, y tiene los valores más bajos en todas las variables, lo cual nos indica que el tiempo de uso de internet que ocupa este grupo en actividades de entretenimiento es extremadamente bajo.

Con la finalidad de poder obtener la incidencia que tiene el ingreso económico con el uso de internet en actividades de entretenimiento, se procedió a aplicar la técnica de regresión logística multinomial, obteniendo como resultado que para todos los rangos de ingresos que tienen los estudiantes de la Pontificia Universidad Católica del Ecuador – Sede Ibarra, indican en un uso **bajo** de internet en actividades de entretenimiento. Lo que nos indica que la hipótesis planteada: “el nivel de ingresos incide sobre el uso de internet en actividades de entretenimiento” para los estudiantes de la Pontificia Universidad Católica del Ecuador – Sede Ibarra es **nula**, toda vez que independientemente del nivel de ingresos que tengan los estudiantes, los mismos presentan un nivel bajo de uso de internet en actividades de entretenimiento.

Estudio emprendido por (Torres, 2012) ha obtenido resultados significativos sobre esta hipótesis aplicada en otras universidades de Ecuador, en donde nos sugiere que los estudiantes que presentan un menor nivel de ingresos tienden a interactuar menos en internet para actividades de entretenimiento, sin embargo en este trabajo de investigación se presenta a esta hipótesis como nula, ya que el 44% de la muestra total de estudiantes presenta un nivel bajo, y el 29% presenta un nivel regular de uso de internet en actividades de entretenimiento, que sumando tenemos que el 73% de estudiantes no tienen un uso significativo de internet en actividades de entretenimiento, lo que hace que todos niveles de ingreso económico de los estudiantes apunten hacia el bajo y regular uso de internet en actividades académicas, y esto nos hace concluir que el ingreso económico de los estudiantes de la Pontificia Universidad Católica del Ecuador – Sede Ibarra, no incide en el uso de internet en actividades de entretenimiento por parte de los mismos.

### **7.3. Incidencia del uso de internet en actividades académicas frente al rendimiento académico.**

La discusión que se presenta en esta sección está alineada a los resultados obtenidos relacionados con la hipótesis 3, misma que sostiene que el uso de internet en actividades académicas incide en el rendimiento académico.

Con la finalidad de poder obtener la incidencia que tiene el uso de internet en actividades académicas con el rendimiento académico, se procedió a aplicar la técnica de regresión logística multinomial, obteniendo como resultado que todos los grupos formados respecto al uso de internet en actividades académicas (alto, medio, regular y bajo) de los estudiantes de la Pontificia Universidad Católica del Ecuador – Sede Ibarra, indican en la aprobación de todas las materias matriculadas. Esto nos indica que la hipótesis planteada: “el uso de internet en actividades académicas incide en el éxito académico de los estudiantes de nivel superior”, para los estudiantes de la Pontificia Universidad Católica del Ecuador – Sede Ibarra es **nula**, toda vez que independientemente del uso del internet en actividades académicas que tengan los estudiantes, estos presentan como aprobadas todas las materias en las que se matricularon.

Estudios emprendidos por (Torres, 2012) y (Castaño Muñoz, 2011) han obtenido resultados positivos sobre esta hipótesis, sin embargo en este trabajo de investigación se presenta a esta hipótesis como nula, ya que el 52% de la muestra total de estudiantes presenta un nivel bajo de uso de internet en actividades académicas, y el 78% de la muestra total de estudiantes presenta su rendimiento académico óptimo (todas las materias aprobadas), lo cual nos hace concluir que el uso de internet en actividades académicas no incide en el éxito académico de los estudiantes de la Pontificia Universidad Católica del Ecuador – Sede Ibarra.

#### **7.4. Incidencia del uso de internet en actividades de entretenimiento frente al rendimiento académico.**

La discusión que se presenta en esta sección está alineada a los resultados obtenidos relacionados con la hipótesis 4, misma que sostiene que el uso de internet en actividades de entretenimiento incide en el rendimiento académico.

Con la finalidad de poder obtener la incidencia que tiene el uso de internet en actividades de entretenimiento con el rendimiento académico, se procedió a aplicar la técnica de regresión logística multinomial, obteniendo como resultado que todos los grupos formados respecto al uso de internet en actividades de entretenimiento (descargas, web social, juegos virtuales, visualizador de videos, medio, regular y bajo) de los estudiantes de la Pontificia Universidad Católica del Ecuador – Sede Ibarra, indican en la aprobación de todas las materias matriculadas. Esto nos indica que la

hipótesis planteada: “el uso de internet en actividades de entretenimiento incide en el éxito académico de los estudiantes de nivel superior”, para los estudiantes de la Pontificia Universidad Católica del Ecuador – Sede Ibarra es **nula**, toda vez que independientemente del uso del internet en actividades de entretenimiento que tengan los estudiantes, estos presentan como aprobadas todas las materias en las que se matricularon.

Estudio emprendido por (Torres, 2012) ha obtenido resultados positivos y el emprendido por (Frangos et al., 2010) ha obtenido resultados negativos sobre esta hipótesis, sin embargo en este trabajo de investigación se presenta a esta hipótesis como **nula**, ya que el 44% de la muestra total de estudiantes presenta un nivel bajo de uso de internet en actividades de entretenimiento, y el 78% de la muestra total de estudiantes presenta su rendimiento académico óptimo (todas las materias aprobadas), lo cual nos hace concluir que el uso de internet en actividades de entretenimiento no incide en el éxito académico de los estudiantes de la Pontificia Universidad Católica del Ecuador – Sede Ibarra.

## **7.5. Conclusiones**

De acuerdo a las encuestas levantadas y al análisis realizado a través de Minería de Datos, el 52% de estudiantes de la Pontificia Universidad Católica del Ecuador presentan un bajo nivel de uso de internet en actividades académicas y el 73% presentan un bajo nivel de uso de internet en actividades de entretenimiento; estos estudiantes están clasificados en los niveles de ingreso analizados, con ello las hipótesis 1 y 2 planteadas se presentan en el resultado como nulas. Esto debido a que el porcentaje de estudiantes que presentan un nivel bajo en uso de internet hacen presencia mayoritaria en cada uno de los niveles de ingreso económico, demostrando con esto que el ingreso económico de los estudiantes no incide en el uso de internet de los mismos.

Como se lo menciono en el párrafo anterior, el porcentaje de estudiantes que presentan nivel bajo en uso de internet supera el 50% del total de la muestra, y por otro lado el 78% de estudiantes presentan un rendimiento académico óptimo (cero materias reprobadas), con ello las hipótesis 3 y 4 planteadas se presentan en el resultado como nulas. Esto debido a que el porcentaje de estudiantes que presentan un nivel bajo en uso de internet supera el 50% al igual que el rendimiento académico

optimo, demostrando con esto que el uso de internet tanto en actividades académicas como para entretenimiento no incide en el rendimiento académico del estudiante.

## **RECOMENDACIONES**

## **8. Recomendaciones**

El uso de una metodología al momento de realizar análisis utilizando minería de datos es esencial ya que obliga a trabajar de forma esquematizada y va dando las pautas para ir avanzando sin dejar de considerar pasos importantes dentro del proceso de la minería de datos.

Los resultados que arrojo esta investigación fueron nulos toda vez que el nivel de uso del internet por parte de los estudiantes es bajo, ante ello es interesante que se amplíe la investigación hacia docentes y varias universidades para poder tener mayor información para analizar todas las tendencias de los estudiantes respecto al uso del internet.

En las encuestas levantadas existen preguntas formuladas tomando varios lapsos de tiempos (semanal, mensual), lo cual hace que algunos encuestados proporcionen información errada, por el lapso de tiempo variable que está entre las preguntas.

En esta investigación el rendimiento académico se midió respecto a materias reprobadas, sin embargo sería importante tomar para esta variable las calificaciones de los estudiantes, lo cual permitirá que esta variable pueda analizarse de una manera más precisa frente a los factores que inciden en la variación de la misma.

Algunas Universidades en Ecuador son reacias a permitir el levantamiento de información y a compartirla por no sentirse preparadas, por lo que sería recomendable que para el emprendimiento de estos trabajos de investigación exista un compromiso de fidelidad de la información entre la institución investigadora y las instituciones analizadas, lo cual facilitaría el proyecto y se obtendría información más precisa.

Como agregado al alcance de esta investigación sería recomendable incluir en la misma la incidencia que está provocando la convergencia digital frente al desempeño de los estudiantes y profesores en la universidad, y la aplicación de la misma la universidad moderna.

## **BIBLIOGRAFIA**

- A4AI. (2013). *Promoviendo la Esequilibridad para un desarrollo incluyente*. Retrieved from <http://a4ai.org/>
- Adell, J. (2011). Entornos personales de aprendizaje (PLE). Retrieved from [www.humanodigital.com.ar/entornos-personales-de-aprendizaje-ple/](http://www.humanodigital.com.ar/entornos-personales-de-aprendizaje-ple/)
- Aluja, T. (2001). La Minería de Datos, entre la Estadística y la Inteligencia Artificial. *Qüestiió*, 25(3), 479–498.
- Arias, A. V., Cabanach, R. G., Núñez Pérez, J. C., & González-Pienda, J. A. (1998). Variables cognitivo-motivacionales, enfoques de aprendizaje y rendimiento académico. *Psicothema*, 10(2), 393–412.
- Ayala, C. (2007). Relación entre el Uso de Internet y el Logro Académico.
- Azevedo, A., & Santos, M. F. (2008). KDD, SEMMA and CRISP-DM: a parallel overview. *IADIS European Conference Data Mining*, 182–185. Retrieved from <http://recipp.ipp.pt/handle/10400.22/136>
- Barolome, A., & Willem, C. (2008). Integración y desarrollos de nuevos elementos de la sintaxis audiovisual en los clips de vídeo digital distribuidos por Internet. Retrieved from <http://www.lmi.ub.es/personal/bartolome/articuloshtml/08viddiginternet.pdf>
- Bruque, D. S. (2013). *El país alcanza las 500 publicaciones científicas relevantes en el 2013*. Retrieved from <http://www.telegrafo.com.ec/sociedad/item/el-pais-alcanza-las-500-publicaciones-cientificas-relevantes-en-2013.html>
- Castaño Muñoz, J. (2011). *EL USO DE INTERNET PARA LA INTERACCIÓN EN EL APRENDIZAJE*. Universitat Oberta de Catalunya.
- Castaño-Muñoz, J. (2010). La desigualdad digital entre los alumnos universitarios de los países desarrollados y su relación con el rendimiento académico. *Revista de La Universidad Y La Sociedad Del Conocimiento*, 7, 1–11.
- CEAACES. (2012). *Resolución cierre de Universidades de categoría E*.
- Coll, C., & Martí, E. (2007). *La educación escolar ante las nuevas tecnologías de la información y la comunicación. Psicología de la educación escolar (Vol. 2)*.

- Retrieved from [http://www.uhu.es/36102/trabajos\\_alumnos/caso\\_10\\_11/\\_private/coll.pdf](http://www.uhu.es/36102/trabajos_alumnos/caso_10_11/_private/coll.pdf) \n [http://books.google.com.co/books?id=uCpHrZCwuZwC&dq=Psicología+de+la+educación+escolar&hl=es-419&sa=X&ei=7dd2Uc\\_EPI289gTXwIDABw&ved=0CDcQ6AEwAQ](http://books.google.com.co/books?id=uCpHrZCwuZwC&dq=Psicología+de+la+educación+escolar&hl=es-419&sa=X&ei=7dd2Uc_EPI289gTXwIDABw&ved=0CDcQ6AEwAQ)
- Cooper, J. (2006). The digital divide: The special case of gender. *Journal of Computer Assisted Learning*, 22(5), 320–334. doi:10.1111/j.1365-2729.2006.00185.x
- Cusí, F. (2003). La alfabetización digital como factor de inclusión social. *La Sociedad de La Información En El Siglo XXI: Un Requisito Para El Desarrollo*. Retrieved from <http://www.desarrollosi.org/PDF/Cont3aPon3.PDF>
- De, U.-U. I., & Telecomunicaciones. (2015). revolution and remaining gaps. *UIT*. Retrieved from [https://www.itu.int/net/pressoffice/press\\_releases/2015/17-es.aspx](https://www.itu.int/net/pressoffice/press_releases/2015/17-es.aspx)
- Diario el Ciudadano. (2014). *Solialización del Código de Economía social del conocimiento en la Asamblea Nacional*. Retrieved from <http://www.elciudadano.gob.ec/senescyt-socializa-codigo-de-economia-social-del-conocimiento-en-la-asamblea-nacional/>
- DiMaggio, P., & Hargittai, E. (2001). From the “Digital Divide” to “Digital Inequality”: Studying Internet use as Penetration Increases. *Center for Arts and Cultural Policy Studies, Princeton University*, 15, 1–23. Retrieved from [http://www.maximise-ict.co.uk/WP15\\_DiMaggioHargittai.pdf](http://www.maximise-ict.co.uk/WP15_DiMaggioHargittai.pdf)
- Duart, J. M., Gil, M., Pujol, M., & Castaño, J. (2008). *universidad en la sociedad red. Usos de Internet en Educación Superior*. Cataluña. Retrieved from [https://www.google.com.ec/url?sa=t&rct=j&q=&esrc=s&source=web&cd=6&cad=rja&uact=8&ved=0CDYQFjAFahUKEwix-sK4vJHJAhWFepAKHVfMDNk&url=http://www.biblioteca.org.ar/libros/141675.pdf&usg=AFQjCNGJCKxgdzNAV\\_F7ZCH6WNPm\\_M6cGg&sig2=m-oibYXb0PregqatrdQVAA](https://www.google.com.ec/url?sa=t&rct=j&q=&esrc=s&source=web&cd=6&cad=rja&uact=8&ved=0CDYQFjAFahUKEwix-sK4vJHJAhWFepAKHVfMDNk&url=http://www.biblioteca.org.ar/libros/141675.pdf&usg=AFQjCNGJCKxgdzNAV_F7ZCH6WNPm_M6cGg&sig2=m-oibYXb0PregqatrdQVAA)
- Dutta, S. (2009). *The Global Information Technology Report 2008–2009 Mobility in a Networked World. Social Computing and Behavioral Modeling*. doi:10.1007/978-1-4419-0056-2\_1

- Francisco Amat, A. (2011). Usando la Web 2.0 para informarse e informar: una experiencia en educación superior. *Teoría de La Educación: Educación Y Cultura En La Sociedad de La Información*, 12, 145–166. Retrieved from [http://campus.usal.es/~revistas\\_trabajo/index.php/revistatesi/article/view/7827/7854](http://campus.usal.es/~revistas_trabajo/index.php/revistatesi/article/view/7827/7854)
- Frangos, C., Frangos, C., & Kiohos, A. (2010). Internet Addiction among Greek University Students: Demographic Associations with the Phenomenon, Using the Greek Version of Young's Internet Addiction Test. *International Journal of Economic Sciences and Applied Research*, 3(1), 49–74. Retrieved from <http://discovery.ucl.ac.uk/20369/>
- Gallardo, A. R. (2006). *la brecha digital y sus determinantes*. Retrieved from [http://132.248.242.3/~publica/archivos/libros/brecha\\_digital\\_y\\_determinantes.pdf](http://132.248.242.3/~publica/archivos/libros/brecha_digital_y_determinantes.pdf)
- Gallardo, J. A. (n.d.). CRISP-DM Metodología para el Desarrollo de Proyectos de Minería de Datos.
- García García, F. (2012). De la convergencia tecnológica a la convergencia comunicativa en la educación y el progreso. *Revista ICONO14. Revista Científica de Comunicación Y Tecnologías Emergentes*, 4(1). doi:10.7195/ri14.v4i1.394
- Gil-juarez, A., Vitores, A., Feliu, J., & Vall-Ilovera, M. (2011). Brecha digital de género: una revisión y una propuesta = Gender digital divide: a review and a proposition. *Teoría de La Educación Y Cultura En La Sociedad de La Información*, 12, 25–53.
- González, P. (2007). *Aprendizaje evolutivo de reglas difusas para descripción de subgrupos*. Universidad de Granada. Retrieved from <file:///C:/Users/Usuario/Downloads/113.pdf>
- Guerrero, I., & Kalman, J. (2010). La inserción de la tecnología en el aula: estabilidad y procesos instituyentes en la práctica docente. *Revista Brasileira de Educação*, 15, 213–230. doi:10.1590/S1413-24782010000200002
- Guerrero, K., & Caballero, D. (2013). El docente-prosumidor y el uso crítico de la web 2.0 en la educación superior. *Revista Sophia: Revista de Investigaciones En Educación*, 9, 79–94. Retrieved from <http://revistas.ugca.edu.co/index.php/sophia/article/view/60>

- Hargittai, E. (2010). Digital Na(t)ives? Variation in internet skills and uses among members of the “net Generation.” *Sociological Inquiry*, 80(1), 92–113. doi:10.1111/j.1475-682X.2009.00317.x
- Hernández Orallo, J., Ramírez Quintana, M. J., & Ferri Ramírez, C. (2004). *Introducción a la Minería de Datos. Introduccion a la mineria de datos.*
- Hinojo, M., & Fernández, A. (2012). El aprendizaje semipresencial o virtual: nueva metodología de aprendizaje en Educación. *Revista Latinoamericana de Ciencias Sociales, Niñez Y Juventud*, 10(1), 159–167. doi:10.3916/C39-2012-03-06
- Howard, P. N., & Massanari, A. (2007). Learning to search and searching to learn: Income, education, and experience online. *Journal of Computer-Mediated Communication*, 12(3), 846–865. doi:10.1111/j.1083-6101.2007.00353.x
- Index, N. R., & Nri, N. R. I. (2012). *Global Information Technology Report 2012. Insight Report.* Retrieved from [http://www3.weforum.org/docs/Global\\_IT\\_Report\\_2012.pdf](http://www3.weforum.org/docs/Global_IT_Report_2012.pdf)
- INEC. (2013). *Tecnologías de la Información y Comunicaciones (TIC´S). Estadísticas Sociales INEC 2013.* doi:10.1109/ColombianCC.2013.6637543
- Instituto Nacional de Astrofísica, Ó. y E. (2012). K-means. Retrieved from <https://ccc.inaoep.mx/~emorales/Cursos/NvoAprend/node78.html>
- Instituto Nacional de Estadísticas y Censos. (2012). *Tecnologías de la Información y Comunicaciones ( TIC ´ S ) 2012 Contenido.* Retrieved from [http://www.inec.gob.ec/sitio\\_tics2012/presentacion.pdf](http://www.inec.gob.ec/sitio_tics2012/presentacion.pdf)
- Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8), 651–666. doi:10.1016/j.patrec.2009.09.011
- Leiner, B. M., Cerf, V. G., Clark, D. D., Kahn, R. E., Kleinrock, L., Lynch, D. C., ... Wolf, S. (2009). A Brief History of the Internet. *ACM SIGCOMM Computer Communication Review*, 39(5), 22–31. doi:10.1145/1629607.1629613
- Lemus, J. A. C. (2014). La evolucion de las tecnologías de la información y las

- comunicaciones en la educación. Retrieved from <http://es.slideshare.net/BekaSom/ja-contreras-lemus-act-18-linea-del-tiempo>
- Lideres, R. (2012). *La convergencia reduce los costos de acceso a tecnología*. Retrieved from <http://www.revistalideres.ec/lideres/convergencia-reduce-costos-acceso-tecnologia.html>
- López, E. (2007). INFLUENCIA DE LA TECNOLOGÍA DE LA INFORMACIÓN EN EL ROL DEL PROFESORADO Y EN LOS PROCESOS DE ENSEÑANZA-APRENDIZAJE. *RIED. Revista Iberoamericana de Educación a Distancia*, 10, 51–60.
- Lopez, M. H., Gonzalez-barrera, A., & Patten, E. (2013). Closing the Digital Divide : Latinos and Technology Adoption, (202). Retrieved from [http://www.pewhispanic.org/files/2013/03/Latinos\\_Social\\_Media\\_and\\_Mobile\\_Tech\\_03-2013\\_final.pdf](http://www.pewhispanic.org/files/2013/03/Latinos_Social_Media_and_Mobile_Tech_03-2013_final.pdf)
- Marín, V. (2011). El alumnado universitario cordobés y la plataforma virtual MOODLE. *PixelBit. Revista de Medios Y Educación*, (38), 121–128. Retrieved from <http://redalyc.uaemex.mx/src/inicio/ArtPdfRed.jsp?iCve=36816200009>
- Marqués Graells, P. (2012). Impacto de las TIC en la educación: Funciones y Limitaciones. *Revista de Investigación 3 Ciencias.*, 10–12.
- Moore, a. (2001). K-means and Hierarchical Clustering. *Statistical Data Mining Tutorials*, 1–24. Retrieved from <http://www.cs.cmu.edu/afs/cs/user/awm/web/tutorials/kmeans11.pdf>
- Morales, C., Soto, S., & Martínez, C. (2005). Estado actual de la aplicación de la minería de datos a los sistemas de enseñanza basada en web. *III Taller Nacional de Minería de Datos*, 49–56. Retrieved from <http://www.lsi.us.es/redmidas/CEDI/papers/189.pdf>
- Morales, E., & Escalante, H. (2015). Aprendizaje Computacional. Retrieved from <https://ccc.inaoep.mx/~emorales/Cursos/NvoAprend/principal.html>
- Morquecho, Erika Priscila Machasilla Piñol, D. S., & Zeas, O. E. U. (2009). *Analisis Del Impacto De La Educacion Gratuita*. ESPOL.

- Ncr, P. C., Spss, J. C., Ncr, R. K., Spss, T. K., Daimlerchrysler, T. R., Spss, C. S., & Daimlerchrysler, R. W. (2000). Step-by-step data mining guide. *SPSS Inc*, 78, 1–78. Retrieved from <http://www.crisp-dm.org/CRISPWP-0800.pdf>
- Norris, P. (2001). *Digital Divide: Civic Engagement, Information Poverty, and the Internet Worldwide. Communication society and politics* (Vol. 21). doi:10.1177/0894439302238974
- Pascual, D., Pla, F., & Sánchez, S. (2007). Algoritmos de agrupamiento. Retrieved from [http://marmota.dlsi.uji.es/WebBIB/papers/2007/1\\_Pascual-MIA-2007.pdf](http://marmota.dlsi.uji.es/WebBIB/papers/2007/1_Pascual-MIA-2007.pdf)
- Peña, I., Córcoles, C. P., & Casado, C. (2006). El profesor 2.0: docencia e investigación desde la Red. *UOC Papers: Revista Sobre La Sociedad Del Conocimiento*, (3), 6. Retrieved from <http://dialnet.unirioja.es/servlet/articulo?codigo=2100874&info=resumen&idioma=SPA>
- Pereira, R. T. (2009). Detección de Patrones de Bajo Rendimiento Académico y Deserción Estudiantil con Técnicas de Minería de Datos. *liisorg*. Retrieved from <http://www.iiis.org/CDs2009/CD2009CSC/CISCI2009/PapersPdf/C692YV.pdf>
- Pick, J. B., & Azari, R. (2008). Global digital divide: Influence of socioeconomic, governmental, and accessibility factors on information technology. *Information Technology for Development*, 14(2), 91–115. doi:10.1002/itdj.20095
- Presidencia de la Republica. (2010). Ley Organica de Educacion Superior en Ecuador. Retrieved from <http://www.ces.gob.ec/descargas/ley-organica-de-educacion-superior>
- Presidencia-Ecuador. (2013). *Gobierno Ecuatoriano Potencia el Desarrollo Tecnológico con Rusia*. Retrieved from <http://www.presidencia.gob.ec/gobierno-ecuatoriano-potenciara-el-desarrollo-tecnologico-con-rusia/>
- PUCE-SI. (2008). Plan estratégico 2009 - 2016 - PUCE-SI. Retrieved from [http://www.pucesi.edu.ec/web/?page\\_id=239](http://www.pucesi.edu.ec/web/?page_id=239)
- Real Academia Española. (2001). *Diccionario de la Lengua Española. Diccionario de la*. Retrieved from <http://lema.rae.es/drae/?val= practica>

- Riquelme, J. C., Ruiz, R., & Gilbert, K. (2006). Minería de datos: Conceptos y tendencias. *Inteligencia Artificial*, 10(29), 11–18.
- Rittberger, M., & Blees, I. (2009). Entorno de aprendizaje de la Web 2.0: Concepto, aplicación y evaluación. *Elearningpapers*, No. 15, 1–20. Retrieved from <http://www.elearningeuropa.info/files/media/media20164.pdf>
- Rodríguez, D., Cuadrado, J., & Sicilia, M. (2007). Comparación de diferentes algoritmos de clustering en la estimación de coste en el desarrollo de software. *Del Software*, 3(1), 6–22. Retrieved from <http://en.scientificcommons.org/44226406>
- Rojas, E. F., Poveda, L., & CEPAL. (2015). *Estado de la banda ancha en America Latina y el Caribe*. Retrieved from <http://www.cepal.org/es/publicaciones/38605-estado-de-la-banda-ancha-en-america-latina-y-el-caribe-2015>
- Romero, C., Romero, C., Ventura, S., Ventura, S., de Castro, C., de Castro, C., ... García, E. (2005). Algoritmos Evolutivos para Descubrimiento de Reglas de Predicción en la Mejora de Sistemas Educativos Adaptativos basados en Web. *Revista Iberoamericana de Informática Educativa*, 2(Julio-Diciembre), 47–60.
- Romero, C., Ventura, S., & Castro, C. De. (2003). Aplicación de algoritmos evolutivos como técnica de minería de datos para la mejora de cursos Hipermedia adaptativos Basados en Web. *Revista Iberoamericana de Educación a Distancia*, 6(2), 141–163. doi:10.5944/ried.6.2.2626
- Senescyt & Inec. (2014). Principales Indicadores de Actividades de Ciencia, Tecnología e Innovación, 32. Retrieved from [http://www.ecuadorencifras.gob.ec/documentos/web-inec/Estadisticas\\_Economicas/Ciencia\\_Tecnologia/Presentacion\\_de\\_principales\\_resultados\\_ACTI.pdf](http://www.ecuadorencifras.gob.ec/documentos/web-inec/Estadisticas_Economicas/Ciencia_Tecnologia/Presentacion_de_principales_resultados_ACTI.pdf)
- Soto Arango, D. E., Mesa Jiménez, F. Y., & Orlando Caro, E. (2012). Convergência digital na universidade colombiana: rumo ao século XXI. *Revista Historia de La Educación Latinoamericana*. Retrieved from <http://dialnet.unirioja.es/servlet/articulo?codigo=4378807&info=resumen&idioma=POR>

- Suría, R. (2012). Redes sociales online y su utilización para mejorar las habilidades sociales en jóvenes con discapacidad. *Escritos de Psicología*, 5(3), 16–23. doi:10.5231/psy.writ.2012.1809
- Tapia Ermel. (2012). Rendimiento Academico.
- Torres e Infante. (2011). Desigualdad digital en la universidad: usos de Internet en Ecuador. *Comunicar*, (37), 81–88. doi:10.3916/C37-2011-02-08
- Torres, J. C. (2012). *Análisis de las relaciones entre los niveles de ingreso, edad y género de los estudiantes, los usos de internet y el rendimiento académico de un grupo de universidades ecuatorianas presenciales.*
- UNESCO, I. de la. (2005). *Hacia las sociedades del conocimiento*. Editorial UNESCO. doi:ISBN 92-3-304000-3
- UTPL. (2013). Publicaciones científicas. Retrieved from <http://investigacion.utpl.edu.ec/es/cifras>
- UTPL. (2015). *UTPL entre las cinco Universidades con mayor producción científica en el Ecuador*. Retrieved from <http://www.utpl.edu.ec/comunicacion/utpl-entre-las-5-universidades-con-mayor-produccion-cientifica-del-pais/#more-43625>
- VALERO, A. T. (2005). *Extracción de Información con Algoritmos de Clasificación*. INAOE. Retrieved from [https://ccc.inaoep.mx/~mmontesg/tesis\\_estudiantes/TesisMaestria-AlbertoTellez.pdf](https://ccc.inaoep.mx/~mmontesg/tesis_estudiantes/TesisMaestria-AlbertoTellez.pdf)
- WEF. (2014). Global Information Technology Report 2014 - El Foro Económico Mundial. Retrieved from <http://reports.weforum.org/global-information-technology-report-2014/>
- Zhang, G.-Q., Zhang, G.-Q., Yang, Q.-F., Cheng, S.-Q., & Zhou, T. (2008). Evolution of the Internet and its cores. *New Journal of Physics*, 10(12), 123027. doi:10.1088/1367-2630/10/12/123027

## **ANEXOS**

## ANEXO 1

### Encuesta a aplicar a los estudiantes de la PUCE-SI

**Estimado estudiante, solicitamos su colaboración contestando esta encuesta, la que permitirá desarrollar una investigación para conocer el uso de internet en las universidades del Ecuador.**

<b>1. Responda la siguiente pregunta</b>	
¿En qué universidad estudia?	

<b>2. Responda la siguiente pregunta</b>	
¿Qué carrera estudia?	

<b>3. Responda la siguiente pregunta</b>	
¿Cuál es su edad?	

<b>4. Responda la siguiente pregunta</b>	<b>Hombre</b>	<b>Mujer</b>
¿Cuál es su género?	( )	( )

<b>5. Los ingresos mensuales de su familia son de:</b>	
Hasta 350 dólares	( )
Hasta 600 dólares	( )
Hasta 1.000 dólares	( )
Hasta 1.500 dólares	( )
Más de 1.500 dólares	( )

<b>6. ¿Desde dónde se conecta habitualmente a Internet? (escoja solo una opción)</b>	
Desde la casa	( )
Desde un cyber café	( )

Desde el trabajo	( )
Desde la Universidad	( )
Desde una red móvil (movistar, claro, cnt)	( )

<b>7. Responda la siguiente pregunta</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>
De 1 a 7, ¿cuántos días a la semana se conecta Internet?	( )	( )	( )	( )	( )	( )	( )

<b>9. Responda las siguientes preguntas</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10</b>
De 1 a 10 su nivel de conocimientos en el manejo de Internet es:	( )	( )	( )	( )	( )	( )	( )	( )	( )	( )

<b>10. Responda las siguientes preguntas</b>	
¿Aproximadamente cuántas horas se conecta cada día?	( ____ )
¿Hace cuántos años se conecta a Internet?	( ____ )

<b>11. En lo referente a las asignaturas en las que está matriculado</b>	
¿Cuántas veces a la semana ingresa a la plataforma virtual de su universidad?	( ____ )
¿Aproximadamente cuántas consultas les hace a sus profesores cada mes?	( ____ )
¿Aproximadamente cuántas consultas realiza a sus compañeros por mes?	( ____ )
¿Aproximadamente cuántos recursos educativos descarga de la plataforma virtual cada mes?	( ____ )
¿Aproximadamente cuántos videos académicos mira en youtube cada mes?	( ____ )
¿Aproximadamente en cuántos foros virtuales participa cada mes?	( ____ )
¿Aproximadamente cuántos post o tweets sobre temas académicos realiza en las redes sociales por mes?	( ____ )
¿Aproximadamente cuántas horas chatea sobre temas académicos cada mes?	( ____ )
¿Aproximadamente cuántas horas busca información académica en internet cada mes?	( ____ )
¿Aproximadamente cuántas horas utiliza la biblioteca virtual de la universidad cada mes?	( ____ )

<b>12. En lo referente al entretenimiento y diversión en internet</b>	
¿Aproximadamente cuántas horas a la semana chatea por diversión?	( ____ )
¿Aproximadamente cuántas horas a la semana utiliza redes sociales?	( ____ )
¿Aproximadamente cuántas horas a la semana utiliza juegos en línea?	( ____ )

¿Aproximadamente cuántas horas a la semana descarga música, videos y programas?	( ____ )
¿Aproximadamente cuántos videos para entretenimiento mira en youtube cada semana?	( ____ )

<b>13. Responda las siguientes preguntas</b>	
¿ Aproximadamente cuántos seguidores tiene en twitter?	( ____ )
¿ Cuántos amigos tiene en facebook?	( ____ )
¿ Cuántos contactos tiene en LinkedIn?	( ____ )

<b>14. Responda con una X en SI o NO a las siguientes preguntas</b>	<b>SI</b>	<b>No</b>
Tiene un blog	( )	( )
Tiene cuenta en youtube	( )	( )
Tiene cuenta en <a href="http://www.del.icio.us">www.del.icio.us</a>	( )	( )
	( )	( )

<b>15. ¿Cuál es su nivel de uso de los siguientes dispositivos? (1 significa no usar y 10 significa utilizar al máximo)</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10</b>
Smartphone con cámara fotográfica y acceso a internet	0	0	0	0	0	0	0	0	0	0
Teléfono móvil con acceso a internet	0	0	0	0	0	0	0	0	0	0
Teléfono móvil sin acceso a internet	0	0	0	0	0	0	0	0	0	0
Computador portátil	0	0	0	0	0	0	0	0	0	0
Tablet (iPad; Galaxy Tab, Kindle, etc)	0	0	0	0	0	0	0	0	0	0
Cámara digital	0	0	0	0	0	0	0	0	0	0
iPod / MP3 Player	0	0	0	0	0	0	0	0	0	0

<b>16. De 1 a 10 valore los siguientes aspectos (1 significa no estar de acuerdo y 10 estar completamente de acuerdo)</b>										
	1	2	3	4	5	6	7	8	9	10

Internet le permite elaborar los trabajos más rápido y con menos esfuerzo	<input type="radio"/>									
Usted confía en la información de internet para realizar sus tareas	<input type="radio"/>									
Internet le permite prescindir de la Biblioteca	<input type="radio"/>									
Internet facilita el proceso de aprendizaje	<input type="radio"/>									
Internet le permite mejorar sus calificaciones	<input type="radio"/>									
Usted presenta trabajos académicos copiados desde Internet	<input type="radio"/>									

<b>17. Responda las siguientes preguntas referentes a sus profesores. (Se recomienda evaluar de forma general a todos sus profesores)</b>	<b>SI</b>	<b>NO</b>	<b>A veces</b>
Su profesor ingresa a la plataforma virtual	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Contesta sus consultas por correo electrónico	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Chatea con usted eventualmente sobre aspectos académicos	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Su profesor comenta en redes sociales sobre temas académicos	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Le sube materiales digitales a la plataforma virtual	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Le recomienda recursos digitales de la biblioteca virtual	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Le recomienda videos sobre temas académicos	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Le plantea cuestionarios o evaluaciones en la plataforma virtual	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Le plantea foros virtuales	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Su profesor tiene una página web, blog o perfil de facebook	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Su profesor tiene cuenta de twitter	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

<b>18. Responda las siguientes preguntas:</b>	
En el semestre anterior, ¿en cuántas asignaturas se matriculó?	( )
En el semestre anterior ¿cuántas asignaturas aprobó?	( )

## ANEXO 2

### Entrevista a aplicar a un directivo(a) de la PUCE-SI

#### 1. Instalaciones físicas

¿Cuenta con salas de cómputo para los estudiantes?

¿Cuenta con Internet inalámbrico para los estudiantes?

¿Ancho de banda total?

¿Número de estudiantes?

¿Tiene software para la gestión académica y administrativa?

Desarrollo propio \_\_\_\_\_

Sistema comercial \_\_\_\_\_

Cuenta con un campus virtual (entorno virtual de aprendizaje)?

Desarrollo propio: \_\_\_\_\_

Sistema comercial: \_\_\_\_\_ Nombre: \_\_\_\_\_

No tiene \_\_\_\_\_

¿Los estudiantes se pueden matricular en línea?

¿Los estudiantes pueden consultar sus calificaciones en línea?

¿Los estudiantes y profesores tienen acceso a una biblioteca virtual?

¿Los profesores suben las calificaciones a Internet?

¿Qué porcentaje de profesores cuenta con un computador personal?

#### 2. Políticas de uso de tecnología

¿Los docentes utilizan obligatoriamente elementos tecnológicos en sus actividades de enseñanza?

Si\_\_\_\_\_ NO\_\_\_\_\_

¿Cuenta la institución con un plan de formación de profesores en el uso de tecnologías para la educación?

¿Aproximadamente qué porcentaje de sus profesores está capacitado en temas tecnológicos?

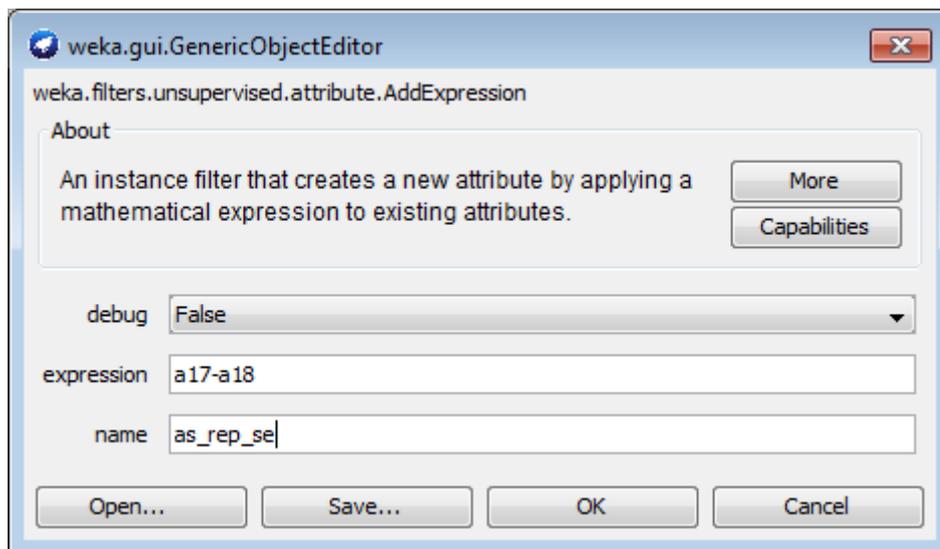
¿Aproximadamente qué porcentaje de sus profesores está capacitado en temas pedagógicos?

¿Tienen algún tipo de incentivo para los profesores que utilizan las tic en sus actividades académicas?

## ANEXO 3

### Proceso de creación de atributo as\_rep\_se (asignaturas reprobadas)

1. Parámetros para la creación del atributo



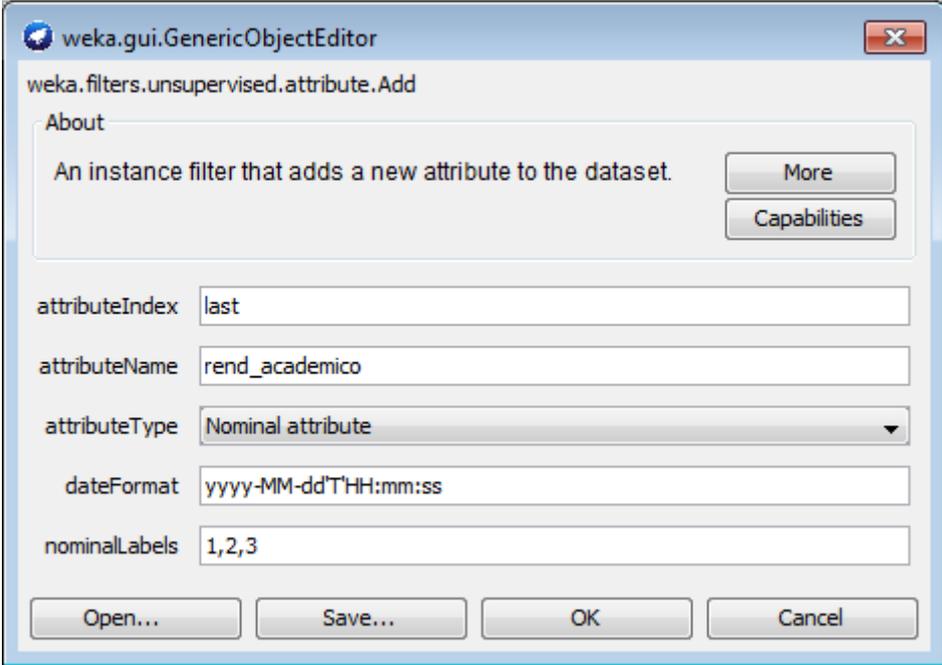
2. Operaciones estadísticas del nuevo atributo

Selected attribute	
Name: as_rep_se	Type: Numeric
Missing: 0 (0%)	Distinct: 6
	Unique: 0 (0%)
Statistic	Value
Minimum	0
Maximum	6
Mean	0.373
StdDev	0.895

## ANEXO 4

### Proceso de crear el nuevo atributo `rend_academico`, discretizando de atributo `as_rep_se`

1. Parámetros ingresados para aplicar el método *Add*



weka.gui.GenericObjectEditor

weka.filters.unsupervised.attribute.Add

About

An instance filter that adds a new attribute to the dataset. [More](#)  
[Capabilities](#)

attributeIndex: last

attributeName: rend\_academico

attributeType: Nominal attribute

dateFormat: yyyy-MM-dd'T'HH:mm:ss

nominalLabels: 1,2,3

Open... Save... OK Cancel

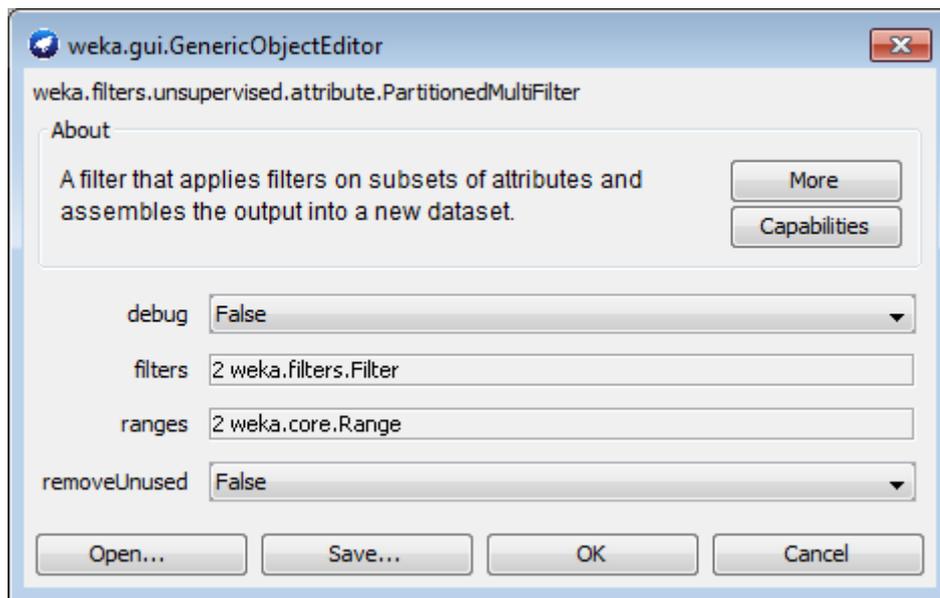
2. Operaciones estadísticas del nuevo atributo

Selected attribute		
Name: rendimiento_academico		Type: Nominal
Missing: 0 (0%)		Distinct: 4
		Unique: 0 (0%)
No.	Label	Count
1	1	386
2	2	72
3	3	24
4	4	16

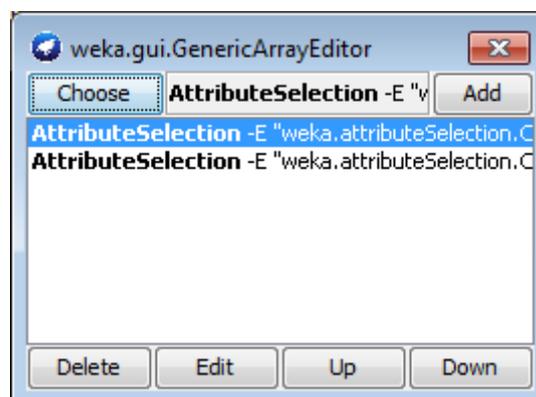
## ANEXO 5

### Proceso de aplicación de análisis factorial a los Dataset “Uso de internet en actividades académicas” y “Uso de internet en actividades de entretenimiento”

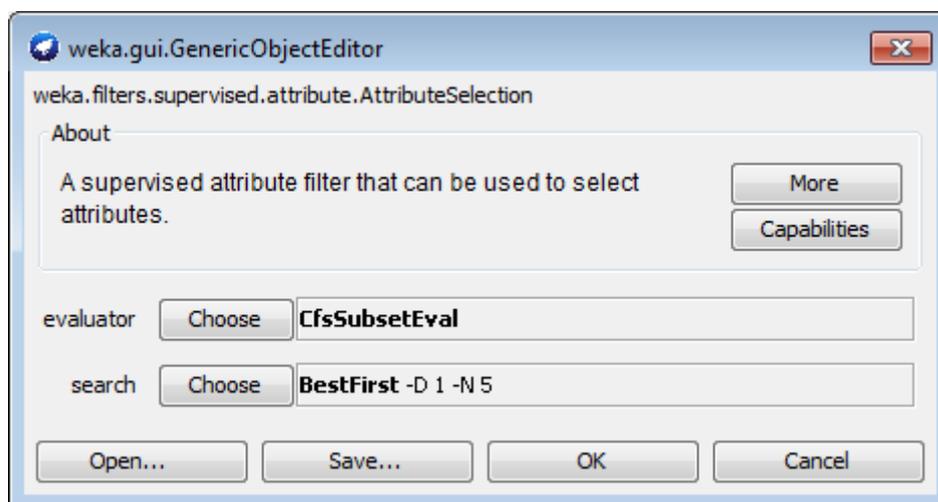
#### 1. Parámetros ingresados para aplicar análisis factorial



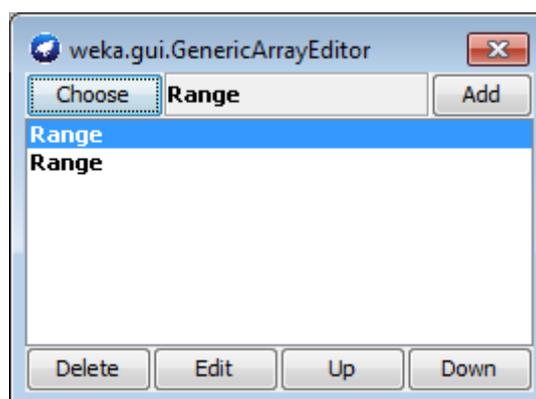
#### 1.1. Parametrización de filtros (se tomo dos filtros, uno para cada DataSet)



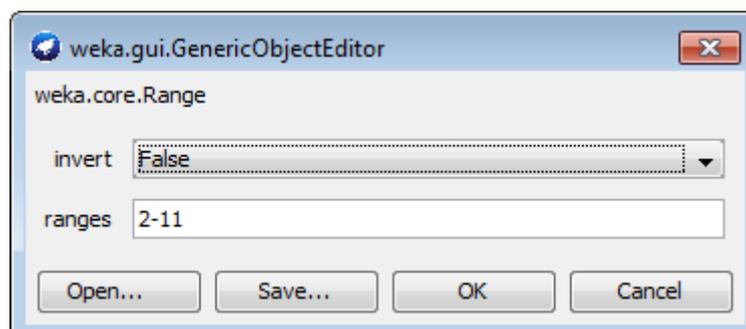
- 1.2. Parametrización de algoritmo AttributeSelection (en este nivel se parametriza el algoritmo evaluador y el de búsqueda).



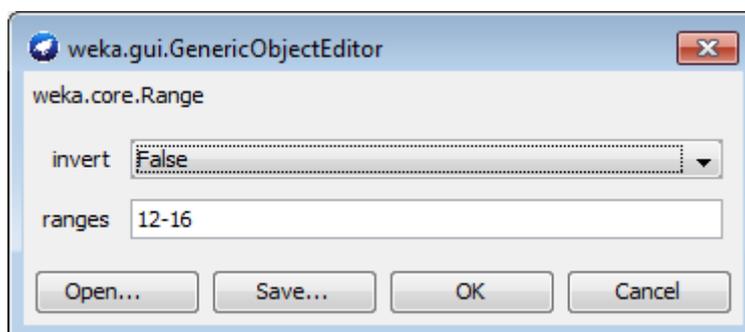
- 1.3. Parametrización de Rangos (se toma dos rango, uno para cada dataset)



- 1.4. Parametrización Rango uno (DataSet "Uso de internet en actividades académicas").



- 1.5. Parametrización Rango dos (DataSet “Uso de internet en actividades de entretenimiento”).



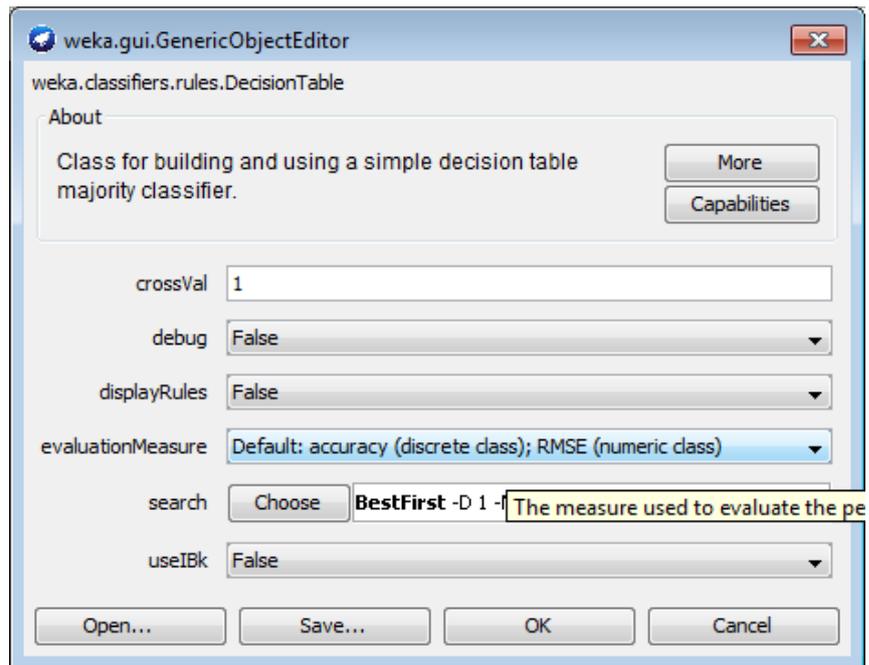
2. Nuevos atributos resultantes luego de aplicar la técnica

No.	Name
1	<input type="checkbox"/> filtered-0-En lo referente a las asignaturas en las que está matriculado - ¿Cuántas veces a la semana ingresa a la pla...
2	<input type="checkbox"/> filtered-0-En lo referente a las asignaturas en las que está matriculado - ¿Aproximadamente cuántos post o tweets s...
3	<input type="checkbox"/> filtered-1-En lo referente al entretenimiento y diversión en internet - ¿Aproximadamente cuántas horas a la semana...
4	<input type="checkbox"/> filtered-1-En lo referente al entretenimiento y diversión en internet - ¿Aproximadamente cuántas horas a la semana ...
5	<input type="checkbox"/> filtered-1-En lo referente al entretenimiento y diversión en internet - ¿Aproximadamente cuántos videos para entret...
6	<input type="checkbox"/> unfiltered-Los ingresos mensuales de su familia son de:
7	<input type="checkbox"/> unfiltered-Responda las siguientes preguntas: - En el semestre anterior, ¿en cuántas asignaturas se matriculó??
8	<input type="checkbox"/> unfiltered-Responda las siguientes preguntas: - En el semestre anterior ¿cuántas asignaturas aprobó??
9	<input type="checkbox"/> as_rep_se

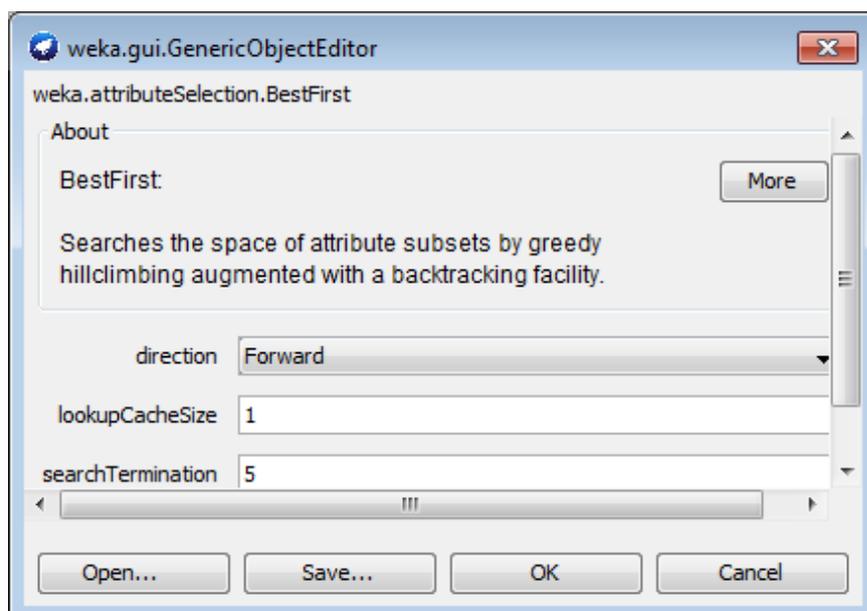
## ANEXO 6

### Proceso de comprobación mediante la aplicación de tabla de decisión

#### 1. Parámetros ingresados para aplicar la tabla de decisión



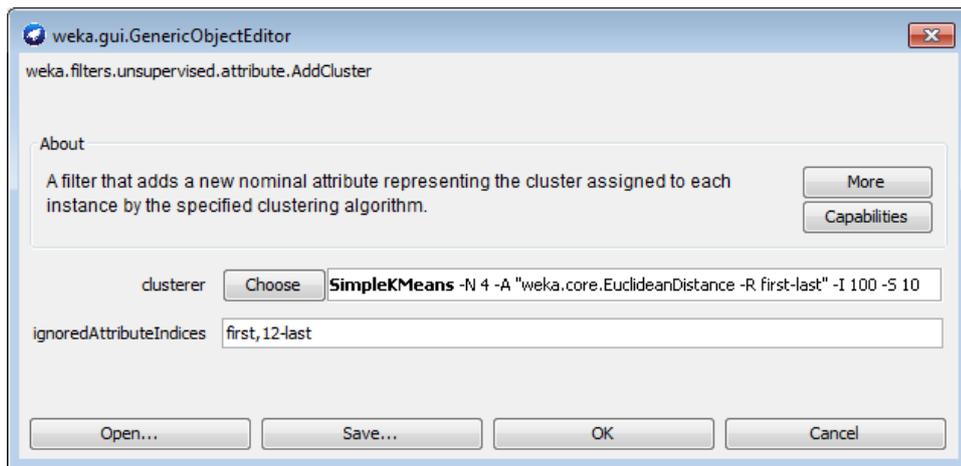
#### 2. Parámetros de método de búsqueda *search*



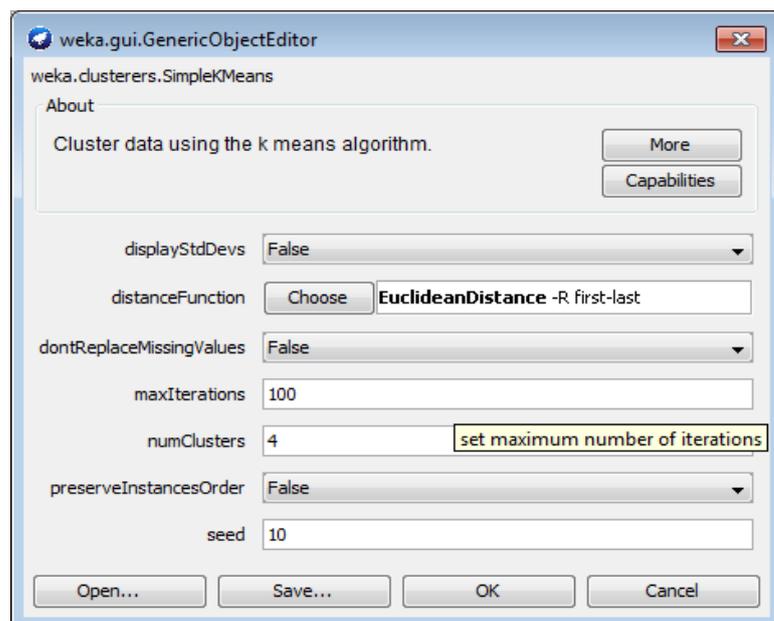
## ANEXO 7

### Proceso de aplicación de clustering a los Dataset “Uso de internet en actividades académicas”.

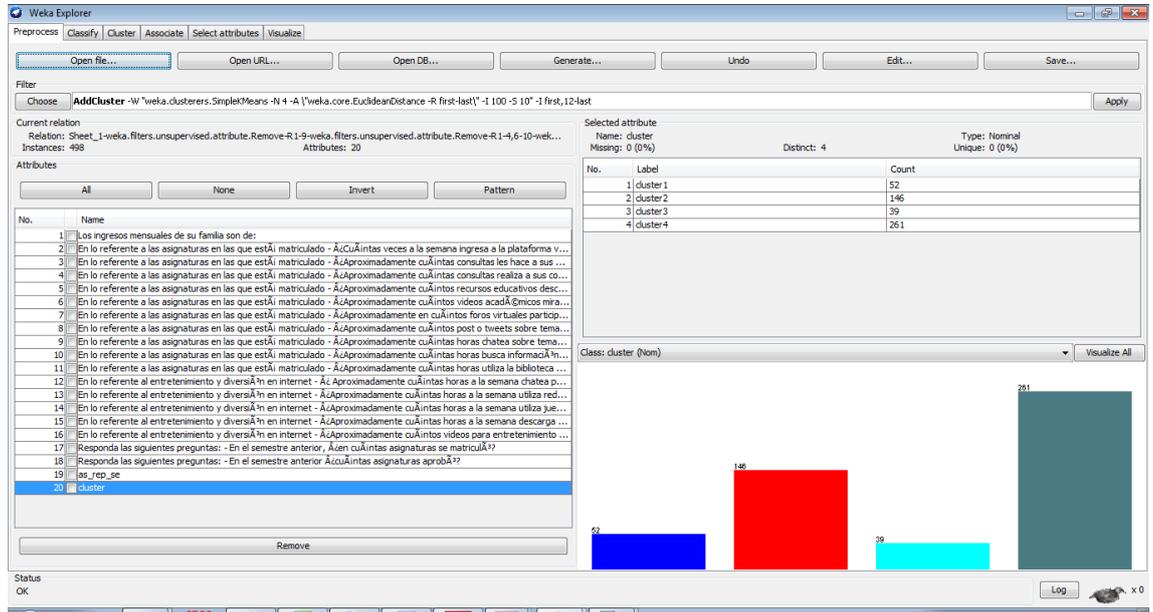
1. Parámetros ingresados para aplicar el método *AddCluster*



2. Parámetros ingresados para aplicar el algoritmo *SimpleKMeans* dentro del método *AddCluster*.



3. Nuevo atributo resultante del proceso realizado, dicha atributo esta con nombre "cluster".



```

Test mode:evaluate on training data

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

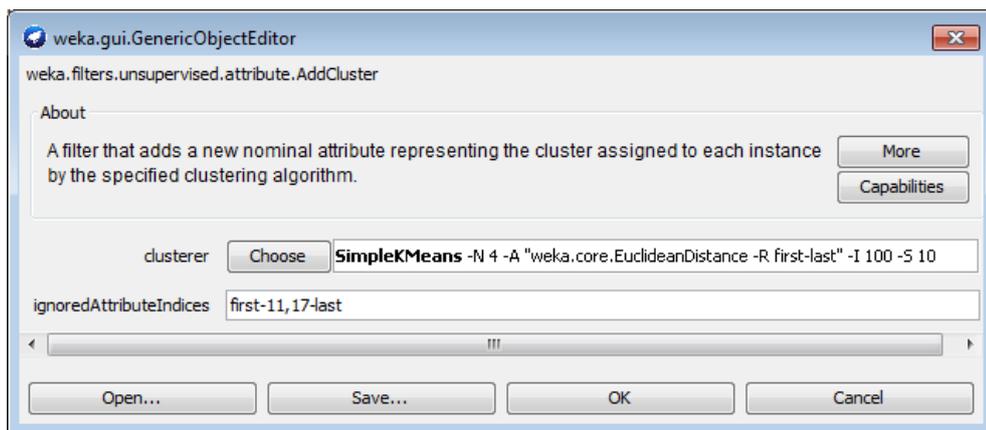
kMeans
=====

Number of iterations: 18
Within cluster sum of squared errors: 164.73941344175094
Missing values globally replaced with mean/mode
  
```

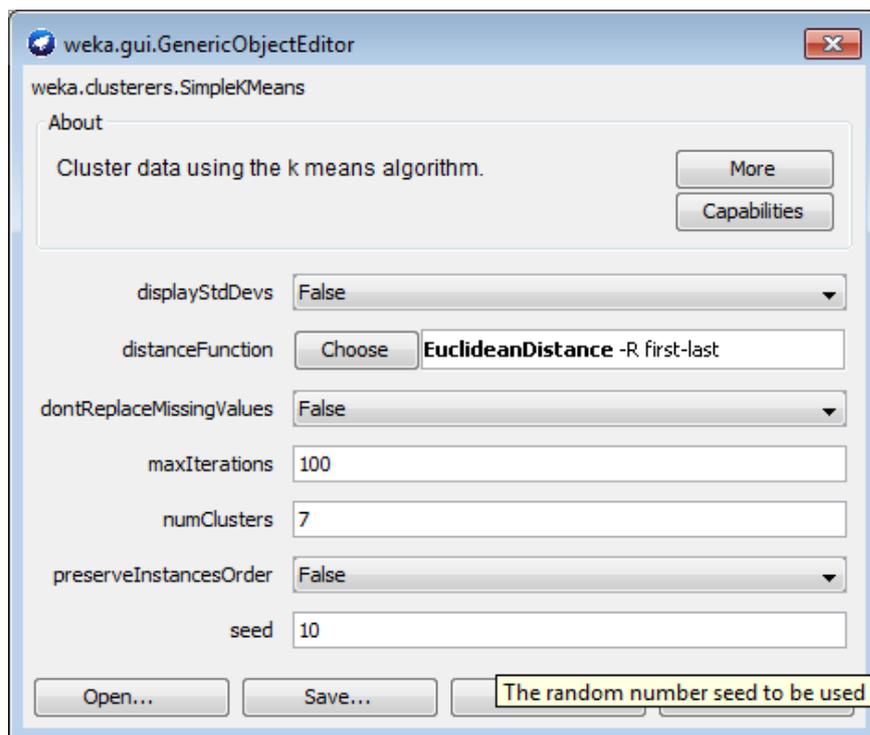
## ANEXO 8

### Proceso de aplicación de clustering a los Dataset “Uso de internet en actividades de entretenimiento”.

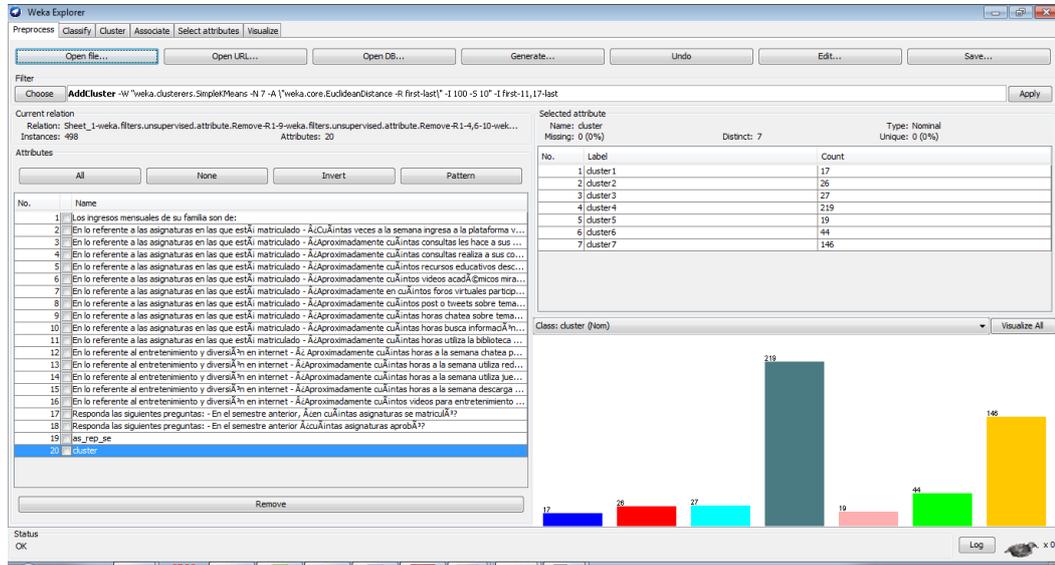
1. Parámetros ingresados para aplicar el método *AddCluster*



2. Parámetros ingresados para aplicar el algoritmo *SimpleKMeans* dentro del método *AddCluster*.



3. Nuevo atributo resultante del proceso realizado, dicha atributo esta con nombre “cluster”.



```

Test mode:evaluate on training data

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

kMeans
=====

Number of iterations: 27
Within cluster sum of squared errors: 34.66354075396136
Missing values globally replaced with mean/mode
    
```