



UNIVERSIDAD TÉCNICA PARTICULAR DE LOJA
La Universidad Católica de Loja

ÁREA TÉCNICA

TÍTULO DE INGENIERO EN INFORMÁTICA

**Análisis de los usos de la tecnología en los estudiantes de la Universidad
Católica Santiago de Guayaquil**

TRABAJO DE TITULACIÓN

AUTOR: Ruiz González, Wilson Armando

DIRECTOR: Cabrera Loayza, María del Carmen, Ing

CENTRO UNIVERSITARIO LOJA

2016

APROBACIÓN DEL DIRECTOR DEL TRABAJO DE TITULACIÓN

Ingeniera.

María del Carmen Cabrera Loayza.

DOCENTE DE LA TITULACIÓN

De mi consideración:

El presente trabajo de titulación: Análisis de los usos de la tecnología en los estudiantes de la Universidad Católica Santiago de Guayaquil, realizado por Ruiz González Wilson Armando, ha sido orientado y revisado durante su ejecución, por cuanto se aprueba la presentación del mismo.

Loja, Enero del 2016

f)

DECLARACIÓN DE AUTORÍA Y CESIÓN DE DERECHOS

“Yo, Ruiz González Wilson Armando declaro ser autor del presente trabajo de fin de titulación: Análisis de los usos de la tecnología en los estudiantes de la Universidad Católica Santiago de Guayaquil, de la Titulación de Ingeniero en Informática, siendo la Ing. María del Carmen Cabrera Loayza directora del presente trabajo; y eximo expresamente a la Universidad Técnica Particular de Loja y a sus representantes legales de posibles reclamos o acciones legales. Además certifico que las ideas, conceptos, procedimientos y resultados vertidos en el presente trabajo investigativo, son de mi exclusiva responsabilidad.

Adicionalmente declaro conocer y aceptar la disposición del Art. 88 del Estatuto Orgánico de la Universidad Técnica Particular de Loja que en su parte pertinente textualmente dice: “Forman parte del patrimonio de la Universidad la propiedad intelectual de investigaciones, trabajos científicos o técnicos y tesis de grado o trabajos de titulación que se realicen con el apoyo financiero, académico o institucional (operativo) de la Universidad”.

f.

Autor: **Ruiz González Wilson Armando**

Cédula: 1103737621

DEDICATORIA

Dedico este trabajo primeramente a Dios quien me ah confortado y ha fortalecido en muchas etapas de mi vida y es gracias a la fe brindada hacia Él que eh podido seguir siempre adelante solventando cualquier inconveniente en mi vida.

De la misma manera quiero dedicar este trabajo a mis padres quienes han estado siempre presentes con sus palabras de aliento para seguir siempre adelante, sin importar lo difícil del camino hacia una meta siempre tuvieron las palabras correctas para levantarme el ánimo y darme fuerzas para seguir.

A mi Hermano quien ha estado presente en todas las etapas de mi vida dándome un consejo.

A mi esposa que ha sabido apoyarme y comprenderme durante la realización de este trabajo y se ha convertido en un pilar fundamental para mi vida.

A mi adorado hijo que con sus locuras, risas y travesuras me inspira a seguir siempre adelante.

AGRADECIMIENTO

Agradezco infinitamente a mis padres por haberme siempre dado el mejor de los ejemplos, la mejor educación y los mejores consejos en las diferentes etapas de mi vida.

A mi amada esposa por su amor, comprensión y apoyo incondicional, por demostrarme que el amor verdadero existe y que todo lo puede superar.

A mi hermano por estar siempre ahí cuando lo necesito.

A mi cuñada Lesly Carvallo por estar presionándome y alentándome para que termine este proyecto.

A mi amiga Jesica Jima por su ayuda y motivación durante el desarrollo de este trabajo.

Un agradecimiento muy especial a mi asesora Ingeniera María del Carmen Cabrera Loayza por su aporte invaluable para la realización de este trabajo.

ÍNDICE DE CONTENIDOS

APROBACIÓN DEL DIRECTOR DEL TRABAJO DE TITULACIÓN	II
DECLARACIÓN DE AUTORÍA Y CESIÓN DE DERECHOS	III
DEDICATORIA	IV
AGRADECIMIENTO	V
INDICE DE CONTENIDOS	VI
ÍNDICE DE TABLAS	X
ÍNDICE DE FIGURAS	XI
INDICE DE FORMULAS	XII
INDICE DE ANEXOS	XII
RESUMEN	1
ABSTRACT	2
INTRODUCCIÓN	3
1.1 Introducción	4
1.2 Problemática	4
1.3 Organización General	6
1.4 Objetivos	7
1.4.1 Objetivo General	7
1.4.2 Objetivos Específicos	7
1.5 Preguntas de Investigación	7
1.6 Hipótesis	7
MARCO TEORICO	8
2.1 Internet	9
2.1.1 Internet como herramienta educativa	9

2.2	E-learning	11
2.3.1	Ventajas y desventajas del E-learning.	11
2.3	Plataformas virtuales de formación	12
2.3.1	Entornos virtuales de aprendizaje	13
2.4	Tecnologías de la información y comunicación	13
2.4.1	Características de las tecnologías de información y comunicación.	14
2.4.2	Importancia de las Tecnologías de Información y comunicación	14
2.4.3	Influencia de las Nuevas Tecnologías en La Educación.	15
2.4.4	Impacto de las TIC en la educación	16
2.5	Gestión del conocimiento	17
2.5.1	Factores clave de éxito en la gestión de conocimiento	17
2.6	Brecha digital	20
2.6.1	Los niveles de la brecha	21
2.7	Técnicas de análisis de datos	22
2.7.1	Minería de datos	22
2.7.1.1	Técnicas de minería de datos	22
2.7.1.2	Herramientas para la aplicación de técnicas de minería de datos	24
2.7.2	Knowledge discovery in databases	26
2.7.2.1	Tipos de datos	26
2.7.2.2	Tipos de modelo	27
2.7.3	Fases del proceso de KDD y minería de datos	27
2.7.4	Minería de datos en la educación	30
2.7.4.1	Aplicaciones de la minería de datos en la educación	30
2.7.4.2	Clasificadores	30
2.7.5	Análisis cluster (clusterización)	31
2.7.5.1	Clúster k-medias	33
2.7.5.2	Clúster jerárquico.	33
2.8	Tablas de Contingencia	34
2.8.1	Prueba Chi-cuadrado.	34
2.8.2	Coeficiente de correlación Bivariada	35
2.8.3	Coeficiente de correlación de Pearson	36
2.8.4	Coeficiente de Determinación	36
2.8.4.1	Coeficiente correlación de rangos de Kendall.	36
2.8.4.1.1	Tau-b- de Kendall:	37
2.8.4.1.2	Tau-c- de Kendall:	37
2.8.5	Análisis Discriminate.	37
2.8.6	Análisis de varianza	38
2.9	Regresión Logística.	39
	METODOLOGÍA	42
3.1	Proceso metodológico	43

3.2	Universo y Tamaño de la muestra de Investigación	44
3.2.1	Universo	44
3.2.2	Tamaño de la muestra de investigación	44
3.3	Técnica de recolección de Información	45
3.4	Obtención de los resultados	46
	RESULTADOS DE LA INVESTIGACIÓN	47
4.1	Resultados	48
4.1.1	Datos Generales de los estudiantes.	48
4.1.2	Hábitos de conexión de la población estudiada	50
4.1.3	Uso de Internet en Actividades Académicas	50
4.1.4	Uso del Internet en actividades de Entretenimiento	54
4.1.5	Uso de Dispositivos Tecnológicos	55
4.1.6	Interacción entre docentes y estudiantes.	56
4.1.7	Rendimiento Académico	57
	ANÁLISIS DE RESULTADOS Y COMPROBACIÓN DE HIPÓTESIS	59
5.1	Relaciones entre variables	60
5.1.1.1	Análisis de la incidencia de la edad sobre resultados obtenidos.	60
5.1.1.2	Análisis de la incidencia del género sobre Variables.	62
5.1.1.3	Análisis de la incidencia de los Ingresos mensuales sobre Variables.	64
5.1	Uso de las tecnologías en actividades académicas	66
5.1.1.4	Reducción de variables	66
5.1.1.5	Clasificación de los estudiantes por el uso de tecnologías en actividades académicas.	67
5.2.3	Resultado.	68
5.3	Uso de las tecnologías en actividades de entretenimiento	69
5.1.1.6	Clasificación de los estudiantes por el uso de las tecnologías en actividades de entretenimiento	69
5.1.1.7	Resultado.	70
5.3	Uso dispositivos electrónicos.	71
5.1.1.8	Clasificación de estudiantes por el uso de dispositivos electrónicos	71
5.1.1.9	Resultado.	72
5.4	Rendimiento académico.	73
5.1.1.10	Clasificación de los estudiantes por su rendimiento académico	73
5.5	Comprobación de hipótesis	74
5.1.1.11	Comprobación de hipótesis 1: El nivel de ingresos determina como se utiliza internet para el aprendizaje	74

5.11.1.12 Comprobación hipótesis 2: El nivel de ingresos determina como se utiliza internet para entretenimiento.	76
5.11.1.13 Comprobación hipótesis 3: El uso de la tecnología en el aprendizaje incide en el rendimiento académico.	78
5.11.1.14 Comprobación hipótesis 4: El uso de la tecnología para entretenimiento incide en el rendimiento académico.	79
OBSERVACIONES FINALES Y CONCLUSIONES	81
6.1 Observaciones finales	82
6.2 Conclusiones generales	85
ANEXOS	86
BIBLIOGRAFÍA	116

ÍNDICE DE TABLAS

Tabla 1. Clasificación de las técnicas de minería de datos	23
Tabla 2. Datos de la edad	49
Tabla 3. Variables que tienen correlación con la edad	61
Tabla 4. Variables que tienen correlación con el género.....	62
Tabla 5. Variables que tienen correlación con los Ingresos mensuales	65
Tabla 6. Porcentaje de precisión de los clúster.....	68
Tabla 7. Porcentaje de precisión de los clúster.....	70
Tabla 8. Porcentaje de precisión de los clúster.....	71
Tabla 9. Rendimiento Académico	74
Tabla 10 Coeficiente modelo de regresión.....	75
Tabla 11. Coeficiente del modelo de regresión	77
Tabla 12. Coeficiente del modelo de regresión	79
Tabla 13. Coeficiente del modelo de regresión	80

ÍNDICE DE FIGURAS

Figura 1. Proceso de KDD y minería de datos	28
Figura 2. Proceso metodológico.....	43
Figura 3. Carreras Universitarias	48
Figura 4. Distribución de estudiantes por los Ingresos mensuales	49
Figura 5. Distribución de estudiantes por el nivel de conocimiento en el manejo de internet	50
Figura 6. Ingreso a la plataforma virtual.....	51
Figura 7. Consultas al profesor	51
Figura 8. Chat temas académicos.....	52
Figura 9. Confianza en la Información de Internet.....	53
Figura 10. Trabajos copiados desde Internet	54
Figura 11. Chat por diversión.....	55
Figura 12. Uso teléfono móvil sin acceso a Internet.....	56
Figura 13. Materias que se matriculo	57
Figura 14. Materias que aprobó	58
Figura 15. Análisis clúster en las actividades académicas	68
Figura 16. Análisis clúster en actividades de entretenimiento	70
Figura 17. Análisis clúster uso de dispositivos electrónicos	72
Figura 18. Rendimiento académico.....	73

ÍNDICE DE FORMULAS

Fórmula 1. Fórmula para calcular chi-cuadrado	34
Fórmula 2. Fórmula del coeficiente de Pearson	36
Fórmula 3. Fórmula para calcular Tau-b	37
Fórmula 4. Fórmula para calcular Tau-c	37
Fórmula 5. Función logística	40
Fórmula 6. Fórmula para poblaciones finitas	44

ÍNDICE DE ANEXOS

ANEXO 1	86
ANEXO 2	89
ANEXO 3	90
ANEXO 4	91
ANEXO 5	92
ANEXO 6	94
ANEXO 7	96
ANEXO 8	98
ANEXO 9	101
ANEXO 10	102
ANEXO 11	104
ANEXO 12	106
ANEXO 13	108
ANEXO 14	109
ANEXO 15	111
ANEXO 16	113
ANEXO 17	115

RESUMEN

Este trabajo de investigación se lo realiza a los estudiantes de la Universidad Católica Santiago de Guayaquil y determina como inciden los ingresos económicos familiares en el uso de internet tanto en actividades académica como de entretenimiento, así también determina como el uso de las tecnologías inciden en el rendimiento académico estudiantil.

Se utiliza el proceso de extracción de conocimiento en base de datos KDD, se emplean técnicas de minería de datos como el análisis clúster para buscar y agrupar estudiantes con características similares en el uso de internet tanto en actividades académicas como de entretenimiento obteniendo de este análisis tres clústers, para comprobar el porcentaje de exactitud de los clúster formados se aplica análisis discriminante.

Para la comprobación de las hipótesis se aplica regresión logística multinomial y binomial, de donde se concluye que los ingresos económicos familiares no determina el uso de internet en actividades académicas ni de entretenimiento; así como el uso de las tecnologías no inciden en el rendimiento académico.

PALABRAS CLAVE: Tecnologías de la Información y Comunicación, Análisis Clúster, KDD, Minería de datos, Regresión Logística

ABSTRACT

This investigation work is done by students at the Catholic University of Guayaquil and determine how they affect the family income in the use of internet both in academic activities and entertainment as well as determined as the use of technology influence student academic performance.

The extraction of knowledge in databases is used, KDD, data mining techniques are used such as cluster analysis to find and group students with similar characteristics in the use of internet in academic activities such as entertainment obtaining this analysis, three clusters, to check the percentage of accuracy of the formed cluster discriminant analysis is applied .

To test the hypothesis multinomial and binomial logistic regression was applied, from which it follows that the family income does not determine the use of internet in academic activities and entertainment; and the use of technologies not affect academic performance.

KEYWORDS: Information Technology and Communication Cluster, KDD , Data Mining , Logistic Regression Analysis.

CAPITULO I
INTRODUCCIÓN

1.1 Introducción

El enorme desarrollo tecnológico ha transformado notablemente la manera de actuar, trabajar, comunicarnos y, por supuesto, de aprender. Así, la sociedad de la información, de la globalización y del comercio on-line, es también la sociedad de la educación virtual, el aprendizaje en línea y del blended learning. Todos estos cambios en el ámbito educativo han sido posibles gracias a tres factores fundamentales: los avances científicos, los adelantos tecnológicos y las demandas sociales según (Herrera Batista, 2009)

Las tecnologías de la Información y Comunicación están en constante desarrollo y por ello la sociedad tiene que adaptarse a este desarrollo para poder explotar al cien por ciento las utilidades que nos brindan en sus diferentes ámbitos de aplicación.

La educación se está direccionando hacia el uso de las nuevas tecnologías para poder llegar de mejor manera hacia los estudiantes. Con este trabajo se quiere determinar si el uso de las tecnologías e internet influye en los estudiantes universitarios de forma positiva o negativa a la hora de realizar sus distintos trabajos académicos, además se va a determinar si al utilizar mayor cantidad de tecnologías en los procesos académicos se mejora el rendimiento académico, también se quiere determinar si el nivel económico familiar influye en la forma de utilizar estas tecnologías tanto en lo académico, entretenimiento e interacción social.

Una vez finalizado ese trabajo se podrá determinar que usos dan a las tecnologías los estudiantes de la Universidad Católica Santiago de Guayaquil, además si el nivel económico familiar influye con la forma de usar las tecnologías tanto en actividades académicas como de entretenimiento y si ese uso del internet influye en el rendimiento académico, así como también que dispositivos tecnológicos son más usados por los estudiantes en sus actividades académicas y de entretenimiento.

1.2 Problemática

Las tecnologías e internet están inmiscuida en varias actividades del ser humano como son la industria, la salud, la educación, etc. (Santillan, 2006) expone que debido a la gran importancia que tiene su presencia en la educación se están imponiendo paulatinamente en cada uno de los procesos académicos.

De este mundo tecnológico y digital nacen varias interrogantes como son: ¿Qué tecnologías influyen en los métodos de enseñanza y aprendizaje?, ¿Qué tecnologías son más eficientes a la hora de llegar a los estudiantes?, ¿Qué tecnologías son de mayor aceptación en ellos a la hora de realizar sus actividades académicas y de entretenimiento?, ¿El uso de internet influye en el rendimiento académico?; cada una de estas interrogantes tenemos que

resolverlas tomando en cuenta la brecha digital existente entre los estudiantes, generada por el nivel económico de los mismos.

Se debe recalcar que aunque la tecnología esté presente en nuestro medio no todos pueden adquirirla e interactuar con ellas o si lo hacen cada persona lo utiliza de acuerdo a sus prioridades no siendo estas necesariamente aplicaciones académicas.

Como lo expone (Espinar Ruiz & González Río, 2008) los jóvenes viven en plena revolución tecnológica, en poco tiempo, el uso de las TIC se ha generalizado hasta tal punto, que la mayoría no concibe la posibilidad de vivir sin ellas, por lo que debemos saber si esta influencia de forma positiva o negativa a la hora de realizar sus actividades académicas, o si la inclusión de estas tecnologías no tiene repercusión en la misma.

Estamos en una sociedad donde el nivel socioeconómico influye en muchos campos, incluidos entre estos la educación, la interacción con las nuevas tecnologías gracias a la facilidad de adquisición de nuevas tecnologías lo que deriva en una mayor habilidad para su uso.

Trabajos de investigación como el de (Naval, Sádaba, & Bringué, 2003) donde se estudia el protagonismo e incidencia de las Tecnologías de la Información y Comunicación (TIC) en la vida y las relaciones sociales de los jóvenes y (Espinar Ruiz & González Río, 2008) donde se analiza, no únicamente el grado de disponibilidad y uso de determinadas tecnologías, sino también, cómo éstas se incorporan a la vida cotidiana de los jóvenes y cómo percibe este colectivo su relación con ellas, exponen la importancia del estudio de las tecnologías y su impacto en la población joven tanto para sus actividades académicas como de interacción social.

Otros trabajos en los que se denota la importancia del estudio de las incidencias y relaciones del uso de internet y variables como el género, la edad y los ingresos económicos han sido realizadas por (Torres, 2012), quien en su investigación demostró que el nivel de ingresos influye en las actividades académicas, así mismo también recomienda investigar si la aparición de lugares de conexión públicos y los bajos costos de las redes de internet logran mitigar esta influencia.

El trabajo de investigación realizado por (Torres Díaz & Infante Moro, 2011) expone la influencia que tiene los ingresos económicos en el uso de internet y la facilidad que tienen quienes cuentan con ingresos económicos altos de adquirir con mayor facilidad medios tecnológicos y además sugieren profundizar en el estudio de la influencia que tiene el género en los usos de internet. Otro trabajo que se toma en cuenta es el de (Castaño-Muñoz, 2010) que pone de manifiesto cómo no solo es importante estudiar las diferencias entre la gente que tiene acceso a Internet y la que no, sino también otras dimensiones como son las habilidades

en el uso de Internet, tiempo en la red y especialmente los tipos de usos que la gente hace de Internet.

1.3 Organización General

El presente proyecto de tesis se organiza en seis capítulos, y cada uno de ellos trata de diferentes temas los que describiremos a continuación:

En el Capítulo I, se expone lo relacionado a la introducción, la problemática que lleva al desarrollo de esta investigación, la organización general de este proyecto, los objetivos generales y específicos a alcanzar, las preguntas de investigación y finalmente hipótesis a comprobarse.

En el Capítulo II se presenta el marco teórico donde se abordan varios temas que son importantes conocer para el desarrollo de este proyecto estos temas son el internet, e-learning, plataformas virtuales de formación, tecnologías de información y comunicación, gestión del conocimiento, brecha digital, técnicas de análisis de datos dentro de la cual se aborda el KDD y cada una de sus fases, tablas de contingencia y finalmente regresión logística.

En el Capítulo III se presenta la metodología utilizada para el desarrollo de este proyecto, aquí se describen el método y su procesos metodológico para el desarrollo de este proyecto, el universo y tamaño de la muestra de investigación, la técnica utilizada para la recolección de la información y como se realizó la obtención de resultados.

En el Capítulo IV se exponen los resultados obtenidos al realizar la investigación a los estudiantes de la UCSG, además se aquí se describen también las incidencias de las variables edad, género e ingresos mensuales sobre las variables obtenidas de la misma y se describen las relaciones entre los estudiantes y el uso de las tecnologías en actividades académicas, de entretenimiento y el uso de dispositivos electrónicos.

En el Capítulo V se presenta el análisis de los resultados y la comprobación de hipótesis por medio de modelos de regresión, donde relacionamos los ingresos mensuales con los tres tipos de clasificaciones obtenidas del análisis clúster del uso de la tecnología en actividades académicas y de entretenimiento.

Por ultimo en el Capítulo VI se trata sobre las discusiones generales y las conclusiones obtenidas durante el desarrollo de este trabajo de titulación así como las recomendaciones que podemos dar basándonos en los resultados obtenidos.

1.4 Objetivos

1.4.1 Objetivo General

Analizar el índice de uso de las tecnologías e internet en las actividades académicas en los estudiantes de la Universidad Católica Santiago de Guayaquil.

1.4.2 Objetivos Específicos

- Determinar la relación entre los niveles socioeconómicos y las actividades académicas y de entretenimiento en los estudiantes de la Universidad Católica Santiago de Guayaquil.
- Determinar la relación entre el nivel ingreso mensuales de los estudiantes y la forma de uso de las tecnologías en actividades académicas y de entretenimiento
- Determinar que tecnologías son más utilizadas por los estudiantes universitarios de la UCSG

1.5 Preguntas de Investigación

Para este proyecto se han planteado las siguientes preguntas de investigación sobre las que se va a basar esta investigación:

- ¿Cómo se relacionan los niveles de ingreso de las familias de los estudiantes universitarios con el uso de internet en actividades académicas y de entretenimiento?
- ¿Cómo se relacionan el rendimiento académico y el uso de internet en actividades académicas y de entretenimiento?

1.6 Hipótesis

Las hipótesis que se plantearon son:

- **Hipótesis 1:** El nivel de ingresos determina como se utiliza internet para el aprendizaje
- **Hipótesis 2:** El nivel de ingresos determina como se utiliza internet para entretenimiento
- **Hipótesis 3:** El uso de la tecnología en el aprendizaje incide en el rendimiento académico
- **Hipótesis 4:** El uso de la tecnología para entretenimiento incide en el rendimiento académico

CAPITULO II
MARCO TEORICO

En este capítulo se desarrolla el marco teórico, es decir, se recopila teoría de los diferentes aspectos que se va a utilizar durante el desarrollo de este proyecto de investigación, además se va a presentar varios postulados y teorías de diferentes autores e investigadores sobre el tema del proyecto.

2.1 Internet

Internet según (Castells, 2013) es la tecnología decisiva de la era de la información, permite producir, distribuir y utilizar información digitalizada en cualquier formato. Esta red global de redes informáticas, que actualmente operan sobre todo a través de plataformas de comunicaciones inalámbricas, nos proporciona la universalidad de una comunicación multimodal e interactiva en cualquier momento y libre de límites espaciales.

En el trabajo realizado por (N. Jiménez, 2008) se menciona a internet como un instrumento que facilita a las personas el rápido acceso a cantidades infinitas de información, a un costo relativamente bajo, sobre cualquier índole y proveniente de cualquier rincón del mundo.

El Internet por sí solo no genera cambios, como expone (Bonilla, 2001) internet inscribe dentro de órdenes y contextos culturales, políticos y sociales, y, por lo general, se convierte en un apéndice o prolongación del poder instituido. En el campo educativo, explica cómo a través de las TIC se reproduce el sistema punitivo de la escuela y se reduce este nuevo lenguaje o herramienta a las formas de control y poder escolar.

Desde diferentes puntos de vista resulta difícil afrontar cualquier aspecto de la sociedad contemporánea sin tomar en consideración el internet. El diario vivir está completamente saturado de tecnologías e internet por lo que ahora se puede distinguir entre dos estados online y offline de acuerdo al aprovechamiento de las tecnologías e internet. De hecho, a menudo observamos que las generaciones jóvenes son incapaces de hablar de internet como una entidad especial ya que esta es parte de su diario vivir como expone (Zelwyn, 2013).

2.1.1 Internet como herramienta educativa

Internet ha sido considerado como una herramienta básicamente educativa para investigadores como (Zelwyn, 2013). De hecho, afirman que las características principales de internet coinciden en gran medida con los intereses centrales de la educación. Por ejemplo, tanto internet como la educación tienen por objeto el intercambio de información, la comunicación y la creación de conocimiento. La naturaleza participativa y comunitaria de las aplicaciones y actividades de internet está en consonancia con las características fundamentales del aprendizaje humano, en concreto, crear, compartir, colaborar y analizar. Así pues, vista la capacidad de internet para hacer posible que estas actividades se lleven a cabo a una escala enorme y de modo casi instantáneo.

Las repercusiones de internet en la educación y el aprendizaje pueden considerarse al menos desde cuatro ángulos diferentes según (Zelwyn, 2013) y estas son:

- Primero se tiene el potencial que brinda a los individuos para aprender con más libertad, sin las ataduras y restricciones del mundo real. Lo que se quiere decir con ello generalmente es que internet reduce las limitaciones locales, espaciales, temporales y geográficas para que los individuos puedan tener acceso a oportunidades de aprendizaje y a medios educativos de alta calidad, con independencia de sus circunstancias particulares. Así pues, internet sería ese medio que permite proporcionar educación en cualquier momento, en cualquier lugar y a cualquier ritmo.
- Segundo, internet es visto como el soporte de una «nueva cultura del aprendizaje» que se basará en los principios «ascendentes» de exploración colectiva, juego e innovación, y no en la instrucción individualizada «descendente» según (Thomas & Brown, 2011). Internet hace posible un aprendizaje de muchos a muchos, en lugar de uno a muchos, desarrollándose en modalidades de aprendizaje y desarrollo cognitivo socio constructivistas de naturaleza profundamente social y cultural. Por ejemplo, se suele afirmar que internet ofrece a los individuos fácil acceso a fuentes de conocimiento teórico y práctico fuera de su entorno inmediato. En este sentido, existe en la actualidad un interés considerable por la capacidad de internet para dar soporte a potentes formas de aprendizaje situado y comunidades de práctica digitalmente dispersas. Así, internet se considera una poderosa herramienta que permite el aprendizaje a través de actividades auténticas y de la interacción entre personas y entornos sociales más amplios como se lo trata en (Luckin, 2010).
- Tercero, existe la impresión que la relación de los individuos con el conocimiento se ha visto alterada por la capacidad que tiene internet para fomentar una conectividad a gran escala entre personas e información. En ocasiones se argumenta que internet favorece formas de creación y consumo de conocimientos que se alejan mucho de las premisas epistemológicas de la enseñanza formal y la instrucción de masas. Las relaciones en red que mantienen los usuarios de internet con la información online han provocado una reevaluación de la naturaleza del proceso de aprendizaje en su conjunto. Desde esta perspectiva, el aprendizaje se entiende como la capacidad de conectarse a nodos y fuentes de información especializada cuando y como se necesite.
- Cuarto, se considera que internet ha personalizado de forma radical el modo en que la gente aprende, de ahí que la educación se convierta en un proceso más individualizado que antes. Internet se asocia con una autonomía y un control social mayor al ofrecer a los individuos más posibilidades de elección en cuanto a la forma y la naturaleza de lo que aprenden, además, de dónde, cuándo y cómo lo hacen además los usuarios de internet

suele atribuírseles una gran capacidad para organizarse y ocuparse personalmente de su propia formación sin necesidad de amoldarse a las normas y expectativas de un sistema educativo.

2.2 E-learning

Según (Molano, Sánchez, & Castillo, 2013) podemos entender a E-learning como un medio electrónico que permite desarrollar procesos de aprendizaje presenciales, a distancia o virtual, logrando una interacción de manera activa con los diferentes agentes involucrados en los procesos de aprendizaje guiados por las tecnologías de la información y la comunicación (TIC); fundamentalmente se apoya en Internet y puede ser descrita como la distribución y el acceso a colecciones coordinadas de materiales de aprendizaje sobre un medio electrónico usando un servidor web para distribuir los materiales, un navegador web para acceder a ellos y los protocolos TCP/IP y HTTP para mediar el intercambio según (Jolliffe, Ritter, & Stevens, 2001).

La característica principal del e-learning es la facilidad que se presta para la comunicación entre los profesores y alumnos apoyada en el uso de la red y de las tecnologías de la Información y comunicación.

2.3.1 Ventajas y desventajas del E-learning.

Entre las principales ventajas según (Almenara, 2006) tenemos:

- Pone a disposición de los alumnos un amplio volumen de información.
- Facilita la actualización de la información y de los contenidos.
- Flexibiliza la información, independientemente del espacio y el tiempo en el cual se encuentren el profesor y el estudiante.
- Permite la deslocalización del conocimiento.
- Facilita la autonomía del estudiante.
- Propicia información precisa en el tiempo correcto.
- Ofrece diferentes herramientas de comunicación sincrónica y asincrónica para los estudiantes y para los profesores.
- Favorece una formación multimedia.
- Facilita una formación grupal y colaborativa.
- Favorece la interactividad en diferentes ámbitos: con la información, con el profesor y entre los alumnos.
- Facilita el uso de los materiales, los objetos de aprendizaje, en diferentes cursos.
- Permite que en los servidores pueda quedar registrada la actividad realizada por los estudiantes.

- Ahorra costos y desplazamiento.

Entre las desventajas tenemos:

- Requiere más inversión de tiempo por parte del profesor.
- Precisa unas mínimas competencias tecnológicas por parte del profesor y de los estudiantes.
- Requiere que los estudiantes tengan habilidades para el aprendizaje autónomo.
- Puede disminuir la calidad de la formación si no se da una relación adecuada profesor-alumno.
- Requiere más trabajo que la convencional.
- Supone la baja calidad de muchos cursos y contenidos actuales.
- Se encuentra con la resistencia al cambio del sistema tradicional.
- Impone soledad y ausencia de referencias físicas.
- Depende de una conexión a Internet, y que ésta sea además rápida.
- Tiene profesorado poco formado.
- Supone problemas de seguridad y además de autenticación por parte del estudiante.
- No hay experiencia en su utilización.
- Existe una brecha digital.

Estas desventajas se solventaran con la capacitación y la disminución de la brecha digital entre los estudiantes y docentes se la lograra mientras más se socialice el uso de las tecnologías de Información y comunicación en todos los procesos educativos.

2.3 Plataformas virtuales de formación

La formación virtual utiliza un softwares específicos denominados genéricamente plataformas de formación virtual. Existen diferentes grupos de entornos de formación según la finalidad de los mismos (Belloch, 2010) expone que ellos son:

- Portales de distribución de contenidos.
- Entornos de trabajo en grupo o de colaboración.
- Sistemas de gestión de Contenidos (Content Management System, CMS).
- Sistemas de gestión del conocimiento (Learning Management System, LMS), también llamados Virtual Learning Enviroment (VLE) o Entornos Virtuales de aprendizaje (EVA).
- Sistemas de gestión de contenidos para el conocimiento o aprendizaje. (Learning Content Management System, LCMS)

2.3.1 Entornos virtuales de aprendizaje

A los entornos virtuales de aprendizaje se los puede describir como medios que facilitan el acceso por medio de navegadores, protegidos generalmente por contraseña o cable de acceso según (Belloch, 2010).

Estos entornos poseen una interface gráfica e intuitiva, además integran de forma coordinada y estructurada los diferentes módulos existentes.

Los EVA según (Belloch, 2010) presentan módulos para la gestión y administración académica, organización de cursos, calendario, materiales digitales, gestión de actividades, seguimiento del estudiante y evaluación del aprendizaje.

Se acoplan a las características y necesidades del usuario, por lo que, disponen de diferentes roles en relación a la actividad que realizan en el EVA. Los niveles de acceso se pueden personalizar y dependen del rol de cada usuario. De tal forma que el EVA es un entorno adaptable a usuarios particulares.

Los EVA permiten la comunicación e interacción bidireccional entre los estudiantes y los profesores o tutores, por lo que se mejora la realización de actividades entre los mismos para facilitar los procesos de aprendizaje.

2.4 Tecnologías de la información y comunicación

Tomando en cuenta la definición de (González et al., 1996), se entiende por "nuevas tecnologías de la información y la comunicación" a los procesos y productos que se derivan de las herramientas de hardware y software, también comprenden los soportes de la información y los canales de comunicación concernientes con el almacenamiento, procesamiento y transmisión digitalizada de la información. Las TIC's permiten obtener y transmitir cualquier tipo de información sea de voz, datos, imágenes, videos etc. de manera rápida y eficaz.

Las tecnologías de la información y comunicación son una parte importante en el desarrollo de la sociedad ya que están inmersas en varios ámbitos de la misma como por ejemplo en lo industrial, académico, cultural, etc. Ya que con estas tecnologías se puede optimizar procesos disminuyendo tiempos de producción y desarrollo y por ende disminuyendo costos en los mismos.

Las tecnologías de la Información y Comunicación permiten enfrentar y solventar de manera eficiente problemas generados con la sobreabundancia de información disponible en la web, en la actualidad se tiene al alcance de la mano un sin número de información sin necesidad de recurrir a los métodos tradicionales de búsqueda como la biblioteca, algún experto o la

experiencia familiar. El conocimiento, sea cual sea el tema, se hace accesible a todos. Ante un buscador, cualquier persona tiene la misma oportunidad de encontrar cualquier contenido que el resto, lo que implica que un conocimiento y manejo eficiente de las TICs puede significar una ventaja ante los demás proporcionando mayores oportunidades laborales y de desarrollo personal.

2.4.1 Características de las tecnologías de información y comunicación.

Como lo expone (Cabero-Almenara, 1994) las características principales son:

- **Inmaterialidad** Al ser su materia prima la información, tiene la posibilidad de construir mensajes sin referentes externos. Mediante la digitalización es posible almacenar grandes cantidades de información, en dispositivos físicos de pequeño tamaño (discos, CD, memorias USB, etc.). A su vez los usuarios pueden acceder a información ubicada en dispositivos electrónicos lejanos, que se transmite utilizando las redes de comunicación, de una forma transparente e inmaterial.
- **Instantaneidad.** Los usuarios pueden acceder a bases y bancos de datos situados en cualquier parte del mundo. Se puede transmitir la información instantáneamente a lugares muy alejados físicamente, mediante las denominadas "autopistas de la información". Se han acuñado términos como ciberespacio, para definir el espacio virtual, no real, en el que se sitúa la información, al no asumir las características físicas del objeto utilizado para su almacenamiento, adquiriendo ese grado de inmediatez e inmaterialidad.
- **Innovación.** Las tecnologías persiguen como objetivo la mejora, el cambio y la superación cualitativa y cuantitativa de su predecesora, y por ende las mejoras de las funciones que estas realizaban, pero las nuevas tecnologías no vienen a superar a sus predecesoras sino más bien las complementan, potencian y revitalizan.

Según (Rosario, 2005) una de las características más importantes de estos entornos es "La interactividad", el uso del ordenador interconectado mediante las redes digitales de comunicación, proporciona una comunicación bidireccional (sincrónica y asincrónica), persona- persona y persona- grupo.

2.4.2 Importancia de las Tecnologías de Información y comunicación

Las tecnologías de la Información y Comunicación han llegado a ser una parte fundamental para los procesos de desarrollo industrial y tecnológico hasta el punto de que ya no es posible diseñar una nueva generación de computadoras sin la asistencia de computadoras para su diseño. Otro ejemplo es la existencia de una maraña de tráfico aéreo, a nivel planetario, la que depende de la asistencia computacional para su coordinación.

Un reporte de (Cerbino & Richero, 2007) expresa que "Ecuador es uno de los países de la región donde más se han realizado esfuerzos entorno a las tecnologías de la información y comunicación" pero es solo hasta el 2012 que el Ecuador elabora su primer documento base

de estándares TIC en concordancia con la propuesta de estándares educativos que llevo a cabo el Ministerio de Educación (ECUADOR, 2010).

Las tecnologías permiten acortar distancias, con ellas podemos comunicarnos con cualquier parte del planeta a través del internet, llamadas telefónicas, videoconferencias etc., lo mejor es que a todas estas tecnologías puede acceder cualquier persona a muy bajo costo. El mundo actual demanda estar siempre conectado ya sea por cuestiones laborales o familiares, lo que hace imprescindible el conocimiento de las mismas su manejo y uso debe ser preocupación del estado pues un país educado en el ámbito tecnológico es un país que va a generar más desarrollo.

A nivel de percepción ciudadana, los cambios producidos por las TIC, se ven, en concordancia, como una fuente de oportunidades y amenazas. Siguiendo los resultados del Informe de Desarrollo Humano en Chile según (Ortega & Guell, 2006), las personas ven a la tecnología como una herramienta que hace más fácil la vida. Esto se da en el acceso a información bancaria o la capacidad de hacer trámites ante el estado sin la necesidad de tener que acudir a una oficina personalmente, evitando la pérdida de tiempo y dinero por transporte y largas esperas. También permite estar informado, el desarrollo de nuevos espacios de socialización y la potenciación de la individualidad.

Con el desarrollo de las tecnologías las personas también adquirieron la facilidad y libertad de buscar la información necesaria para satisfacer sus intereses sin la necesidad de ser un experto en el uso de las mismas, aunque para realizar una búsqueda efectiva hay que tener un conocimiento básico, debido a que en la red, se puede encontrar un sin número de información no necesariamente verificada o de calidad.

2.4.3 Influencia de las Nuevas Tecnologías en La Educación.

Nos encontramos ante una revolución tecnológica; en el que las tecnologías de la información y comunicación son parte de la vida diaria. Estas nuevas tecnologías plantean nuevos paradigmas, revolucionan el mundo de la enseñanza. Po medio de estas tecnologías se pueden visitar museos de ciudades de todo el mundo, leer libros, hacer cursos, aprender idiomas, visitar países, ponerse en contacto con gente de otras culturas, acceder a textos y documentos sin tener que moverse de una silla, etc, a través de Internet.

Las nuevas tecnologías son parte integrante de los procesos de educación por esto cada vez más universidades en el mundo están exigiendo la alfabetización electrónica como uno de los requisitos en sus exámenes de acceso y de graduación, por considerar que es un objetivo esencial preparar a los futuros profesionales para la era digital en los centros de trabajo.

Como lo expone (Rosario, 2005) la mayoría de las instituciones de educación superior cuentan, en mayor o menor medida, con equipos informáticos que posibilitan el acceso a

Internet de los alumnos. Así, los universitarios, incluso aquellos que por problemas económicos no cuentan con computadores en sus hogares, pueden acceder a un mundo que antes era exclusivo de las clases pudientes, teniendo la oportunidad de visitar museos y accediendo a conocimientos disponibles gratuitamente. Es en este sentido, que el papel del profesor universitario es fundamental. Cuanto más se inculque en los universitarios la posibilidad de utilizar las nuevas tecnologías, más amplio será el mundo que obra para ellos y las oportunidades que tengan de encontrar trabajo.

2.4.4 Impacto de las TIC en la educación

Según lo expuesto por (Coll, 2008) el primer y principal argumento sobre el impacto esperado de la incorporación de las TIC a la educación tiene que ver con el papel de estas en la llamada Sociedad de la Información, refiriéndonos al argumento según el cual en el nuevo escenario social, económico, político y cultural de la sociedad de la Información el conocimiento se ha convertido en la mercancía más valiosa de todas, y la educación y la formación en las vías para producirla y adquirirla.

La dificultad de establecer relaciones causales más o menos directas entre la utilización de las TIC y la mejora del aprendizaje ha llevado a algunos autores como (Jonassen, Howland, Moore, & Marra, 2002) y (Twining, 2002) a desplazar el foco de atención hacia el estudio de cómo la incorporación de las TIC a los procesos formales y escolares de enseñanza y aprendizaje pueden modificar, y modifican de hecho en ocasiones, las prácticas educativas.

Al incorporar las TIC a la educación se espera determinar que papeles van a cumplir las mismas dentro de la educación, como van a influenciarla y a donde van a llevarla. Se hace referencia a la manifestación según la cual en el nuevo escenario social, económico, político y cultural de la sociedad de la información el conocimiento es el bien más valioso de todos, y a este conocimiento se lo puede producir u obtener gracias a la educación y los diferentes procesos formativos. La educación no es el único medio para impulsar el desarrollo, la socialización y la culturización de las personas, sino más bien la educación se concibe como un medio para desarrollar la identidad nacional y ciudadanía.

Las TIC en la educación se convierten en una prioridad estratégica para las políticas de desarrollo nacionales. Pero lo que interesa recalcar es el inminente crecimiento del uso de las TIC en los procesos educativos en todos los niveles de la sociedad. Es importante señalar dos puntos uno de ellos es que las tecnologías hacen posible, por medio de la reducción de la brecha digital, que más personas tengan acceso a la formación y la educación. Otro punto a señalar es que gracias a las tecnologías multimedia e Internet, disponemos de nuevos recursos en los entornos educativos.

El uso combinado de las tecnologías multimedia e Internet nos permite el aprendizaje fácilmente en cualquier lugar.

La transformación de los escenarios educativos tradicionales, la incorporación de las TIC a la educación formal y escolar es a menudo justificada, reclamada o promovida, según los casos, con el argumento de su potencial contribución a la mejora del aprendizaje y de la calidad de la enseñanza. Las explicaciones en pro de la implementación de las TIC en todos los niveles de la educación no tienen discusión sino más bien se justifican en las prestaciones que brindan para realizar nuevas metodologías de enseñanza.

Se debe investigar hasta donde y bajo qué parámetros las TIC pueden llegar a cambiar las prácticas educativas. Lo que nos interesa es saber cómo las TIC's mejoran la enseñanza y el aprendizaje y analizar el uso efectivo que le dan los profesores y alumnos durante estos procesos; y por otra, se prevén gracias a las mismas mejoras en el aprendizaje de los alumnos y mayor participación e implicación en estas actividades, En conclusión, según(Carneiro, Toscano, & Díaz, 2009) la verdadera importancia de las TICs está en la forma que estas son utilizadas por profesores y alumnos en los diferentes procesos de enseñanza y el aprendizaje.

2.5 Gestión del conocimiento

La gestión del conocimiento es un concepto relativamente nuevo por lo que aún no existe un cuerpo doctrinal sólido y estructurado, a pesar de toda la importancia que abarca este concepto.

Tras un detenido análisis de las definiciones y las características propias de la creación y gestión del conocimiento, (Rodríguez, 2006) menciona que podemos considerar a la gestión del conocimiento como un conjunto de procesos sistemáticos (identificación y captación del capital intelectual; tratamiento, desarrollo y compartimiento del conocimiento; y su utilización) orientados al desarrollo organizacional y/o personal y, consecuentemente, a la generación de una ventaja competitiva para la organización y/o el individuo.

Como se menciona en (Aja Quiroja, 2002) "La gestión del conocimiento pretende lograr que la organización pueda tener acceso a la información y la facilidad requerida para adaptarla continuamente a los cambios tanto internos como externos del medio ambiente, además expone que una correcta gestión de la información y del conocimiento se deriva de una correcta implementación de la gerencia de la calidad".

2.5.1 Factores clave de éxito en la gestión de conocimiento

Como menciona (Alzate Molina & Gallego Álvarez, 2012) existen 7 factores claves que deben tener en cuenta las organizaciones para crear una adecuada cultura del conocimiento, estos factores son:

a. Procesos

Las organizaciones deben procurar que sus actividades se encuentren en base de procesos y no de personas, se lo logra a través de:

- Conectar gente con gente y gente con conocimiento.
- Conectar gente con información.
- Facilitar conversión de información a conocimiento.
- Encapsular conocimiento para facilitar transferencia.
- Diseminar conocimiento.

b. Estructura organizacional

Se debe procurar que la estructura organizacional de las instituciones no sea de manera vertical si no horizontal, esto se logra con:

- Equipos de trabajo
- Grupos de trabajo
- Comunidades de práctica

c. Gente

Se debe concientizar a los empleados de la organización de la importancia que tienen estos en ella, debido a que son ellos los que conforman una organización, es por esto que es de vital importancia que estos interioricen esto. Adicionalmente se debe velar porque exista un adecuado relacionamiento interpersonal entre estos.

- La gestión de conocimiento con la gente se hace:
- Demostrando que no se crea conocimiento sin gente.
- Con modelos de competencias implantados y utilizados para la gestión cotidiana.
- Con confianza (especialmente la percepción de qué gano si comparto mis conocimientos).
- Promoviendo el trabajo en equipo.
- Generando motivación hacia el conocimiento.

d. Sistema de Reconocimiento

Se debe establecer una política clara de incentivos que motive a las personas a generar y compartir conocimientos. Esto se hace buscando estrategias que generen Gestión del Conocimiento con sistemas de reconocimiento, algunos de estos sistemas son:

- Que se reconozca contribución al conocimiento
- Reconocimiento de que conocimiento está ligado al ego
- Reconocimiento por:

- Nuevas o mejores competencias o conocimiento relevante.
- Asumir nuevos proyectos o responsabilidades
- Desarrollar colegas.
- Destacarse como experto en un tema o competencia (mejor práctica).

e. Sistemas de Información

El éxito de la implementación de la Gestión de Conocimiento radica en poseer un excelente sistemas información, el cual les permita a las personas acceder a esta de una manera rápida y oportuna.

La Gestión del Conocimiento desde los sistemas de información se hace:

- Con el conocimiento capturado, compartido, creado.
- A través de la conversación.
- A través de relaciones estimulantes para la transferencia de conocimiento.

f. Liderazgo

Es de suma importancia el compromiso de la alta gerencia en la implementación de la Gestión del Conocimiento, para esto es necesario que los líderes de la organización, contribuyan con:

- Convertirse en soporte a la gestión de conocimiento.
- Dar ejemplo.
- Generando un alto nivel de confianza en empleados.
- Teniendo la convicción de que conocimiento es apalancador de resultados.

g. Cultura

Este el último factor clave y el más importante, debido a que la organización debe adoptar una cultura que trabaje día a día por fomentar la Gestión o Gerencia del Conocimiento en todos sus procesos y personas.

Para (Rodríguez, 2006) en cambio los factores clave de éxito para una correcta gestión de conocimientos son:

- Cultura orientada al conocimiento:** debe existir una cultura favorable y compatible con la gestión de conocimientos.
- Infraestructura técnica e institucional:** A la gestión de conocimientos se la puede desarrollar de mejor manera si existe una correcta infraestructura tecnológica y el personal ha desarrollado las competencias necesarias para hacer uso de ella.
- Respaldo del personal directivo:** debe existir el completo apoyo del equipo directo para la creación y gestión del conocimiento.

- d. **Vínculo con el valor económico o valor de mercado:** los procesos de gestión del conocimiento pueden resultar muy costosos, por tanto, es necesario que se traduzcan en algún tipo de beneficios para la organización.
- e. **Orientación del proceso:** es aconsejable realizar una buena evaluación diagnóstica que nos oriente el desarrollo del proceso.
- f. **Claridad de objetivo y lenguaje:** es importante estar claros con aquello que queremos conseguir.
- g. **Prácticas de motivación:** el conocimiento es personal por lo que es fundamental motivar e incentivar a los miembros de la organización para que lo compartan, lo usen y lo creen de forma habitual.
- h. **Estructura de conocimiento:** es fundamental la creación de una estructura de conocimiento flexible.
- i. **Múltiples canales para la transferencia de conocimiento:** se debe proporcionar diferentes canales y situaciones que faciliten la transferencia de conocimiento.

2.6 Brecha digital

Diferentes Organismos mundiales entre ellos la CEPAL implantaron el concepto de brecha digital en la zona para referirse al ambiente que se vivió en Latinoamérica, donde se veía una dificultad para la implementación de las TIC respecto a países del Primer Mundo como se lo explica (J. Jiménez, 2010).

La brecha digital es una realidad innegable si ponemos a observar la desigualdad en la capacidad de transmisión y uso de Internet entre las diferentes regiones del planeta.

La brecha digital se refiere según (Álamo, 2003) a la distancia existente entre diferentes áreas y diferentes estratos socio económicos en relación a las ventajas para acceder y obtener las nuevas tecnologías de la información y la comunicación así como también el uso de Internet, lo que al final refleja la desigualdad entre los países y dentro de ellos.

Según estas investigaciones, hubo una infraestructura deficiente, así como escasas iniciativas que exponían que los diversos gobiernos de la región, por ellos mismos no podrían enfrentar este reto.

La brecha digital no se enmarca en solo unos matices sino más bien abarca un sin número de estos como son lo social, económico, tecnológico, uso de tecnologías, acceso a internet etc.

Para estudiar la brecha digital hay que resolver diferentes preguntas como: ¿La forma y facilidad que tienen los jóvenes para acceder al internet y hacer uso de las nuevas tecnologías difiere a la de los adultos mayores?, ¿Existe desigualdad entre el área rural y urbana?, ¿Cómo

la tecnología penetra en cada peldaño de la sociedad? Ante estas preguntas también deben preguntarse los gobiernos que están haciendo ante la inminente llegada de esta era digital, ¿Qué están haciendo para reducir la diferencia en el acceso a internet en los diferentes niveles socio económicos?

2.6.1 Los niveles de la brecha

Si se toma la teoría que la brecha digital es un entorno de una única dimensión, se están cometiendo diversos errores al realizar un análisis inadecuado ante una realidad que se desarticula en varios niveles. En base a esto podemos decir que dentro de cada región existe una brecha digital y esta se repite cuando buscamos a nivel nacional. En este contexto, en todos los países, solo una proporción de la población dispone de los medios tecnológicos más poderosos. La brecha que el porcentaje restante de la población debe afrontar en referencia al uso de las tecnologías es equivalente a las que soporta la población en los países en vías de desarrollo.

La brecha digital, en si es un problema común en el mundo y muestra características similares en diferentes países. Los países desarrollados empezaron a establecer nuevas estrategias y a adoptar varias medidas para solventar el problema en base a nuevas políticas gubernamentales. Los bajos recursos son un obstáculo extremadamente grande para que los países en desarrollo adopten esas estrategias. Se han creado perjuicios generacionales y de género, por lo que muchas investigaciones revelan que los hombres ya dominan la sociedad de la información y comunicación poniendo en evidencia la necesidad de volver neutro este proceso y evitar que las mujeres se queden atrás en el uso de las nuevas tecnologías.

En conclusión estos datos y previsiones que han aparecido solo revelan que la brecha digital existe entre sociedades y dentro de ellas. Por lo que el uso de Internet esta estratificado no solo en el género, sino también es más frecuente en la población joven que en personas mayores, habitantes del medio urbano y rural con personas de niveles socio económicos altos.

Para generalizar se va tomar en cuenta 5 formas generales de desigualdad que son:

- Medios Técnicos (hardware y conexiones) por medios de las cuales las personas acceden a la Web, de aquí la importancia de los equipos adecuados para la conexión y trabajo ya que el uso de equipos inadecuados generaría retrasos y dificultades en el uso de los medios tecnológicos.
- Variedad en la forma y grado en la que las personas practican o se preparan en el uso del internet, donde definimos donde los usuarios se conectan con mayor frecuencia, cuán lejos deben ir para utilizar los mismos.
- Desigualdad en la habilidad para el uso del internet

Donde encontramos cuatro niveles de habilidad para el uso del internet, conocimiento medio, realizar búsquedas, bajar información y un conocimiento de fondo.

- Desigualdad en el soporte social.

Los usuarios por lo general obtienen soporte social de un asistente técnico formal, asistencia de amigos o familiares

- Variedad en el propósito de uso del internet, donde no todos los usuarios utilizan los medios tecnológicos de forma científica sino más bien la buscan como medio de comunicación o entretenimiento.

2.7 Técnicas de análisis de datos

2.7.1 Minería de datos

Existen varias definiciones de minería de datos entre ellas se va a citar las siguientes:

(M. Torrado, 2011) define la minería de datos o Data Mining como la extracción de información desconocida no trivial y potencialmente útil de una gran cantidad de información.

Otro concepto es el dado por (Larrañaga & Inza, 2006) quien dice que la minería de datos es el proceso de obtener conocimiento útil y entendible, anteriormente desconocido, desde grandes conjuntos de datos almacenados en varios formatos.

En la actualidad la minería de datos se utiliza en diferentes ámbitos como el financiero, análisis de Mercado, comercio, educación, procesos industriales, etc.

En si se puede decir que la minería de datos trata en general de obtener información de grandes bases de datos para así poder tomar decisiones en busca de un objetivo.

2.7.1.1 Técnicas de minería de datos

Las técnicas de minería de datos según (Weiss & Indurkha, 1998) se clasifican en dos clases: supervisados o predictivos y no supervisados o de descubrimiento del conocimiento.

Así mismo dicen (Weiss & Indurkha, 1998) que los algoritmos supervisados o predictivos predicen el valor de un atributo etiqueta de un conjunto de datos, conocidos otros atributos (atributos descriptivos). A partir de datos cuya etiqueta se conoce se induce una relación entre dicha etiqueta y otra serie de atributos. Esas relaciones sirven para realizar la predicción en datos cuya etiqueta es desconocida. Esta forma de trabajar se conoce como aprendizaje supervisado y se desarrolla en dos fases: Entrenamiento (construcción de un modelo usando un subconjunto de datos con etiqueta conocida) y prueba (prueba del modelo sobre el resto de los datos).

Cuando una aplicación no es lo suficientemente madura y no tiene el potencial necesario para una solución predictiva hay que recurrir a los métodos no supervisados o de descubrimiento del conocimiento, que descubren patrones y tendencias sin tener ningún tipo de conocimiento

previo acerca de cuáles son los patrones buscados, en los datos actuales (no utilizan datos históricos).

El descubrimiento de esa información sirve para llevar a cabo acciones y obtener un beneficio (científico o de negocio) de ellas, todo el proceso de modelado se lleva a cabo sobre un conjunto de ejemplos formados tan sólo por entradas al sistemas, en estos no se tiene información sobre las categorías de esos ejemplos, por lo que, el sistema tiene que ser capaz de reconocer patrones para poder etiquetar nuevas entradas. En la tabla siguiente se muestran algunas de las técnicas de minería de ambas categorías.

Tabla 1. Clasificación de las técnicas de minería de datos

Supervisados	No Supervisados
Arboles de decisión	Detección de desviaciones
Inducción neuronal	Segmentación
Regresión	Agrupamiento "Clustering"
Series temporales	Reglas de asociación
	Patrones secuenciales

Fuente:(García, Quintales, García, & Polo, 1997)

Otra forma de clasificar las técnicas de minería de datos según (Aranda & Sotolongo, 2013) por su utilidad son:

- **Las técnicas de predicción:** Estas nos permiten obtener pronósticos de comportamientos futuros en base a datos obtenidos con anterioridad, de donde se pueden aplicar frecuentemente.
- **Las técnicas de agrupamiento:** Estas técnicas agrupan datos dentro de un número de géneros preestablecidas o no, están parten de criterios de distancia o similitud, de forma que las clases sean lo más análogas entre sí y diferentes de las otras clases como se menciona (Rodríguez Suárez, 2009).
- **Las técnicas de reglas de asociación:** Estas técnicas permiten definir las posibles relaciones o correlaciones entre los distintos sucesos o acciones supuestamente independientes; Como lo dice (Molina López & García Herrero) se puede reconocer como la ocurrencia de un suceso u acción puede influenciar la aparición de otros.
- **Las técnicas de clasificación:** Estas técnicas especifican una serie de clases, en la que se pueden agrupar los diferentes casos. Dentro de este grupo se encuentran las técnicas de árboles de decisión y reglas de inducción.

2.7.1.2 Herramientas para la aplicación de técnicas de minería de datos

Para poder utilizar las técnicas de minería de datos antes mencionadas hay varias herramientas; entre estas herramientas existen aquellas que son indistintas del sistema gestor de bases de datos y otras se derivan de un gestor de bases de datos específico.

- **Herramientas originarias del gestor**

Como se menciona en (Aranda & Sotolongo, 2013) ORACLE y SQL Server han implementado técnicas para el análisis de datos con el objetivo de facilitar el proceso de descubrimiento de conocimiento para la toma de decisiones.

SQL Server Data Mining: contiene varias características necesarias para crear complejas soluciones de minería de datos, y nos permite:

- Aplicar soluciones de minería de datos utilizando Microsoft Excel.
- Entender cómo, cuándo y dónde aplicar los algoritmos que se incluyen en el servidor de SQL.
- Realizar la extracción de datos de procesamiento analítico en línea (OLAP).
- Utilizar SQL Server Management Studio para acceder y proteger los objetos de minería de datos.
- Utilizar SQL Server Business Intelligence Development Studio para crear y gestionar proyectos de minería de datos.

Hay muchas ventajas de la minería de datos de Microsoft como lo menciona (Aranda & Sotolongo, 2013) estas ventajas son:

- La integración estrecha con la plataforma de base de datos de clase mundial SQL Server, pues aprovecha el desempeño, la seguridad y las características de optimización de SQL Server.
- La extensibilidad, ya que podemos extender la minería de datos de Microsoft para implementar algoritmos que no vienen incluidos en el producto.

Entre los algoritmos implementados por Microsoft podemos mencionar según (Aranda & Sotolongo, 2013) los siguientes:

- Árboles de decisión.
- Bayes naive.
- Clústeres.

- Redes neuronales.
- Serie temporal.
- Regresión lineal.
- Clústeres de secuencia.
- Asociación.

Oracle Data Mining: Como se lo expone en (Haberstroh, 2008) este permite que las empresas puedan desarrollar aplicaciones de inteligencia de negocio avanzadas que utilicen las bases de datos corporativas, revelen nuevos conocimientos e integren esa información en aplicaciones comerciales.

Oracle Data Mining incorpora algunas funcionalidades de minería de datos para realizar clasificaciones, agrupamiento, predicciones y asociaciones entre las que podemos mencionar según (Aranda & Sotolongo, 2013):

- Agrupamiento (k-means, O-Cluster).
- Árboles de decisión.
- Atributo relevante.
- Característica de selección.
- Clasificador bayesiano (naive bayes).
- Máquinas de soporte vectorial (support vector machines).
- Modelos lineales generalizados
- Reglas de asociación (APRIORI).
- **Herramientas independientes del gestor**

Una de las herramientas libres de mayor uso para la minería de datos según (Ramesh, 2011) es Weka (Waikato Environment for Knowledge Analysis) que no es más que una herramienta visual de distribución libre para el análisis y la extracción de conocimiento a partir de datos.

Las principales ventajas de Weka según (Aranda & Sotolongo, 2013) son:

- Es multiplataforma.
- Contiene una extensa colección de técnicas para pre-procesamiento y modelado de datos.
- Es fácil de usar, gracias a su interfaz gráfica.
- Soporta varias tareas de minería de datos, especialmente pre-procesamiento, agrupamiento, clasificación, regresión, visualización y selección.

- Permite combinar varios algoritmos basados en técnicas de minería de datos, para obtener mejores resultados en el descubrimiento de conocimiento.
- Es capaz de mostrar los datos en varios tipos de gráficos con el objetivo de proporcionar una mejor comprensión y un mejor análisis.

2.7.2 Knowledge discovery in databases

Un concepto bastante explícito de KDD, knowledge Discovery in databases, es dado por (Riquelme, Ruiz, & Gilbert, 2006) y nos dice:

“La integración de un conjunto de áreas que tienen como propósito la identificación de un conocimiento obtenido a partir de las bases de datos que aporten información hacia la toma de decisiones.”

KDD tiene por finalidad procesar grandes cantidades de datos automáticamente, identificando los patrones más significativos y relevantes que se convertirán en la información para poder presentarlos como conocimiento apropiado para lograr los objetivos del usuario

2.7.2.1 Tipos de datos

Existen varios tipos de datos sobre los cuales se puede realizar la minería de datos estos pueden ser:

- Bases de datos relacionales: que pueden ser una colección de relaciones (tablas). Tabla como conjunto de atributos (variables, columnas, campos) conteniendo tuplas (casos, filas, registros) Presentación tabular: atributo-valor (vista minable)
- Bases de datos espaciales: Datos geográficos, imágenes médicas, redes de transporte o tráfico.
- Bases de datos temporales: que se presentan en distintos instantes o intervalos de tiempo temporales
- Bases de datos documentales: estas son documentos de texto, variables desde palabras hasta los resúmenes
- Bases de datos multimedia: aquí tenemos a las imágenes, audio, video.
- La Word Wide Web: es el repositorio de información más grande y diverso en la actualidad
 - Minería del contenido: encontrar patrones en las páginas web
 - Minería de la estructura: estudia los hipervínculos y URLs
 - Minería del uso: análisis de la navegación

Los datos deben ser según (Vieira, Ortiz, & Carvajal, 2009) sin errores de medición, coherentes, sin falta de atributos, concernientes al problema y no tener redundancia en ellos.

2.7.2.2 Tipos de modelo

Existen varios tipos de modelos que se puede aplicar como por ejemplo:

- Modelos de datos a conocimiento por medio de modelos computacionales
- Modelos descriptivos: Estos identifican patrones que explican los datos, dentro de estos tenemos:
 - Reglas de asociación: Estos expresan patrones de comportamiento en los datos
 - Clustering: Estos agrupan casos de iguales características.
- Modelos predictivos: estos modelos predicen valores de variables de interés a partir de valores de otras variables llamadas predictoras.
 - Regresión: En esta la variable a predecir es continua
 - Clasificación supervisada: la variable a predecir es discreta y puede ser nominal u ordinal según sea el caso.

2.7.3 Fases del proceso de KDD y minería de datos

Para realizar los procesos de extracción del conocimiento existen varios proceso que varían según cada autor pero en si nos llevan al mismo objetivo por ejemplo (Cabena, Hadjinian, Stadler, Verhees, & Zanasi, 1998) divide al KDD en las siguientes fases:

- a. **Determinación de objetivos**
- b. **Preparación de datos**
 - Selección: Identificación de las fuentes de información externas e internas y selección del subconjunto de datos necesario.
 - Pre procesamiento: estudio de la calidad de los datos y determinación de las operaciones de minería que se pueden realizar.
- c. **Transformación de datos:** conversión de datos en un modelo analítico.
- d. **Minería de datos:** tratamiento automatizado de los datos seleccionados con una combinación apropiada de algoritmos.
- e. **Análisis de resultados:** interpretación de los resultados obtenidos en la etapa anterior, generalmente con la ayuda de una técnica de visualización.
- f. **Asimilación de conocimiento:** aplicación del conocimiento descubierto.

Por otra parte (Larrañaga & Inza, 2006) dicen que los procesos a seguir son los siguientes:

- 1 Integración y recopilación de datos
- 2 Selección, limpieza y transformación de datos
- 3 Minería de datos
- 4 Evaluación e interpretación de resultados
- 5 Difusión y uso

Como se puede ver las fases expuestas por (Larrañaga & Inza, 2006) y (Cabena et al., 1998) son muy similares pero para este proyecto vamos a tomar como referencia los procesos descritos por (Larrañaga & Inza, 2006).

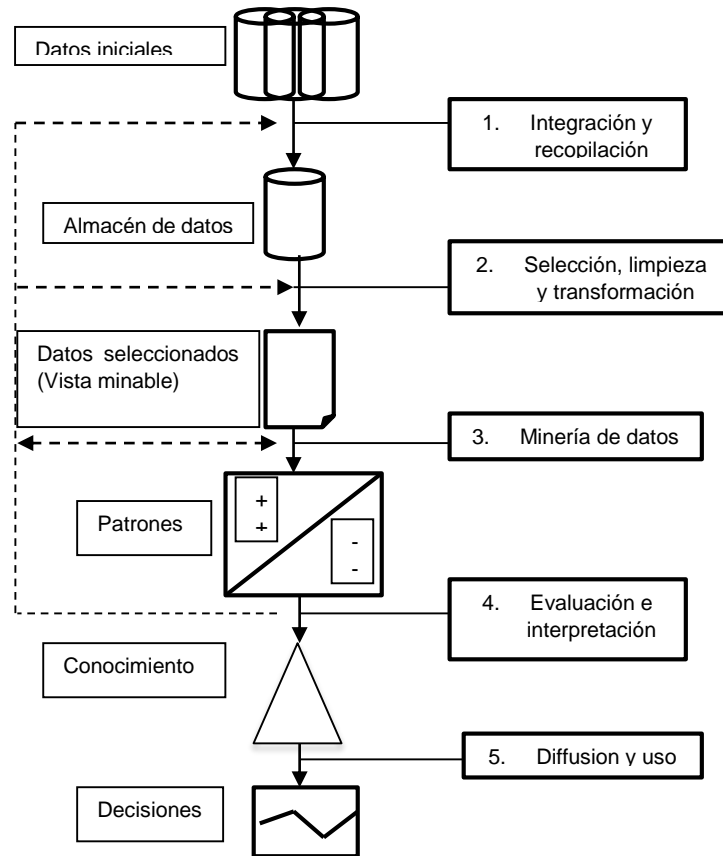


Figura 1. Proceso de KDD y minería de datos

Fuente:(Larrañaga & Inza, 2006)

a. Integración y recopilación de datos

En la Integración y recopilación de datos encontramos lo siguiente:

- Procesamiento transaccional en línea (On-Line Transaction Processing, OLTP): suficiente para necesidades diarias (facturación, control de inventario, ...)
- Decisiones estratégicas basadas en el análisis, la planificación y la predicción: datos en varios departamentos
- Cada fuente de datos distintos formatos de registro, diferentes grados de agregación, diferentes claves primarios.
- Integración de múltiples bases de datos: almacenes de datos (data warehousing)
- Almacén de datos aconsejable cuando el volumen de información es grande. No estrictamente necesario (archivos de texto, hojas de cálculo).

b. Selección, limpieza y transformación

- Calidad del conocimiento descubierto depende (además del algoritmo de minería) de la calidad de los datos analizados
- Presencia de datos que no se ajustan al comportamiento general de los datos (outliers)
- Presencia de datos perdidos (missing values)
- Selección de variables relevantes (feature subset selection)
- Selección de casos aleatoria en bases de datos de tamaño ingente. Muestreo aleatorio simple, por conglomerados, estratificado, polietápico
- Construcción automática de nuevas variables que faciliten el proceso de minería de datos
- Discretización de variables continuas

c. Minería de datos

Dentro de la minería de datos se aplica:

- **Modelos descriptivos**

Reglas de asociación:

Clustering: particional, probabilístico, jerárquico, conceptual

- **Modelos predictivos:**

Regresión: regresión lineal, regression tree, model tree, additive regression

Clasificación supervisada: clasificadores Bayesianos, regresión logística, redes neuronales, arboles de clasificación, inducción de reglas, K-NN, combinación de clasificadores

d. Evaluación e interpretación

Técnicas de evaluación: validación simple (training + test), validación cruzada con k-rodajas, bootstrapping

Reglas de asociación: cobertura (soporte), confianza

Clustering: variabilidad intra y entre

Regresión: error cuadrático medio

Clasificación supervisada: porcentaje de bien clasificados, matriz de confusión, análisis ROC

Modelos precisos, comprensibles (inteligibles) e interesantes (útiles y novedosos)

e. Difusión y uso

Difusión: necesario distribuir, comunicar a los posibles usuarios, integrarlo en el know-how de la organización

Medir la evolución del modelo a lo largo del tiempo (patrones tipo pueden cambiar)

Modelo debe cada cierto tiempo de ser: reevaluado, reentrenado y reconstruido

2.7.4 Minería de datos en la educación

Según (Cristóbal Romero, Ventura, & Hervás, 2005) las técnicas de minería de datos en la educación las podemos analizar desde dos puntos de vista:

- **Minería de Datos Orientada hacia los autores.** La finalidad de esta es ayudar a los educadores autores de los sistemas de e-learning a mejorar su funcionamiento a partir de la información obtenida sobre la utilización de los alumnos. Entre las principales aplicaciones según (Cristóbal Romero et al., 2005) tenemos: alcanzar una mejor y más eficiente realimentación de la enseñanza, conocer como los estudiantes aprenden en el web, evaluándolos por sus patrones en la navegación, otra aplicación es el poder cambiar los contenidos en el sitio web para así personalizar los diferentes cursos.

- **Minería de datos orientado hacia los estudiantes.** De igual manera (Cristóbal Romero et al., 2005) exponen que la finalidad de esta es efectuar recomendaciones a los alumnos durante su interacción con el sistema de e-learning para así mejorar su aprendizaje. Las principales aplicaciones de esta son: proponer mejores experiencias de aprendizaje a los estudiantes, dependiendo del progreso del estudiante poder adaptar los cursos y ayudarlos dándoles sugerencias y recomendar para el mejor aprovechamiento de los mismos.

Aunque la minería de datos aplicada a educación es bastante nueva ya cuenta con un gran número de investigadores, y para muestra de esto existen múltiples contribuciones publicadas en diferentes congresos internacionales (ICCE, ICALT, ITS, Elearn, PAKDD, GECCO, UM, AH, WISE, ISDE, etc.) y revistas (IJEL, IEEE Education, UMUAI, etc.).

El número de congresos e investigaciones sobre la minería de datos en la educación como los expresan (Cristóbal Romero et al., 2005) está incrementándose enormemente año tras año, indicando de esta manera la gran importancia que está teniendo y el interés que ah despertando en la sociedad.

2.7.4.1 Aplicaciones de la minería de datos en la educación

Existen varios componentes que son importantes para la minería de datos en la educación como por ejemplo los clasificadores, agrupamiento y reglas de asociación que nos ayudan a reconocer la información oculta para los diferentes actores en las instituciones educativas. Los principales métodos para la generación de un modelo predictivo dentro de la educación son los clasificadores y clustering según (Ballesteros, Sánchez, & García, 2013).

2.7.4.2 Clasificadores

Los clasificadores utilizan modelos supervisados según (C. Romero, Ventura, Pechenizkiy, & Baker, 2010) clasificar es colocar un objeto o categoría en base a sus características, para la creación de modelos predictivos que describan clases de datos importantes de tendencias

futuras tomando en cuenta que la minería de datos pretende obtener reglas para la división de los datos en una clase o categoría para la mejor comprensión de los mismos.

Lo que se pretende con los clasificadores es enseñar a una máquina una manera de clasificar los objetos a través del análisis de una hipótesis definida en donde las clases son conocidas, asimismo analizar el funcionamiento y el comportamiento de implementar los algoritmos de clasificación sobre un conjunto de ejemplos univariados y multivariados. Según (Ballesteros et al., 2013) las técnicas de clasificación se las realiza de forma diferente dependiendo el método utilizado, a pesar de esto todas ellas parten del mismo conjunto de datos considerado como muestra y todas terminan descubriendo una predicción de acuerdo a un porcentaje de acierto.

Para (Ballesteros et al., 2013) los clasificadores discriminativos determinan un valor de la clase por cada registro del dato, es decir, si M es un clasificador, $C = \{c_1, c_2, \dots, c_i\}$ el conjunto de valores de la clase y " t " un registro del dato, entonces la predicción de la clase es $M(t) = c_i$ para un solo " i " (donde i es el valor del atributo), otra alternativa es un clasificador probabilístico, definiendo la probabilidad de una clase para todas las filas clasificadas, la predicción de la clase es $M(t) = [P(C = c_1|t), \dots, P(C = c_i|t)]$, donde $P(C = c_i|t)$ es la probabilidad de que " t " pertenece a la clase c_i , (C. Romero et al., 2010).

Según (Ballesteros, 2012) en el sector educativo, las técnicas de minería de datos educativa en base a la clasificación y predicción son utilizadas para la comprensión del comportamiento estudiantil, de forma que se pretende analizar el comportamiento que tienen estos en los diversos mecanismos de aprendizaje que hoy en día son utilizados en los diversos sistemas de aprendizaje-enseñanza con el objetivo de generar modelos educativos que puedan analizar e incrementar el nivel participativo y de aprendizaje de los estudiantes en función de modelos predictivos que conlleven a un alto impacto y puedan ser implementados en la actualidad como pueden ser el modelo presencial, modelo virtual y/o modelo mixto.

Los proceso enseñanza-aprendizaje, los sistemas de control escolar así como las consecuencias de implementar nuevas metodologías educativas, nos brindan una cantidad considerable de información, gracias a la cual podemos aplicar de múltiples métodos de clasificación que ofrecen según (Cristóbal Romero et al., 2005) resultados favorables en su combinación entre tiempo de ejecución y porcentaje de acierto.

2.7.5 Análisis cluster (clusterización)

Como menciona (Manzano, 2008) cuando se tiene un conjunto de variables que caracterizan varios individuos, el análisis de conglomerados o clúster nos permite formar grupos de

individuos, de manera que cada grupo formado contenga dentro de si individuos con características similares pero distintas de los otros grupos formados

El análisis Clúster nos permite clasificar sujetos u objetos a partir de características similares estos grupos o clúster se obtienen y configuran una vez estudiadas y analizadas las agrupaciones obtenidas.

Como menciona (Ballesteros et al., 2013) los métodos de agrupamiento a diferencia de la clasificación y la predicción que utilizan modelos supervisados, analizan los registros con etiquetas de clase o categoría los objetos de datos, de tal forma que los métodos de agrupamiento no están interesados en modelar un conjunto de relaciones y un conjunto de respuestas que pertenezcan a cada valor asociado, este método utiliza el modelado no-supervisado donde analiza los objetos de datos sin consultar con una clase conocida, es decir, las etiquetas de la clase no se encuentran en los datos de entrenamiento por el hecho de que el objeto es desconocido, de tal forma que el clustering es utilizado para la generalización de etiquetas, es decir, la búsqueda y construcción de grupos formados de tal forma que los patrones/objetos dentro de un conjunto tienen una alta similitud en comparación entre sí denominada función de medida de asociación, donde cada grupo formado puede ser interpretado como una clase de objetos de los cuales se pueden derivar reglas.

El análisis de conglomerados es similar al análisis factorial, difiere una de otra en que la factorización es un poco flexible en algunos de sus supuestos y siempre estima de la misma manera la matriz de distancias, la aglomeración es menos restrictiva en sus supuestos y admite varios métodos de estimación de la matriz de distancias.

Como técnica de agrupación de casos, el análisis de conglomerados es similar al análisis discriminante, difieren estos en que el análisis discriminante realiza la clasificación tomando como referencia un criterio o variable dependiente, el análisis de conglomerados permite detectar el número óptimo de grupos y su composición únicamente a partir de la similaridad existente entre los casos estos conceptos se los expone en(UCM, 2011a).

Las técnicas de análisis clúster tienen como objetivo principal la clasificación de un cierto número de individuos en unos cuantos conglomerados o clúster. El principio general que guía la formación de los grupos es la minimización de la variación interna y la maximización de las distancias entre los clúster. La partición obtenida depende en gran proporción de la medida de similitud empleada para conocer la separación entre los grupos y del método elegido para la formación de tales grupos.

Existen dos tipos de análisis de conglomerados que son mayormente utilizados y son: el análisis de conglomerados jerárquico y el análisis de conglomerados de K medias. El método

jerárquico es utilizado para determinar el número óptimo de conglomerados existente en los datos y el contenido de los mismos. El método de K medias nos permite procesar un número ilimitado de casos, pero sólo nos permite utilizar un solo método de aglomeración y requiere que se proponga con anterioridad el número de conglomerados que se quiere obtener.

2.7.5.1 Clúster k-medias

Se utiliza Clúster K-medias cuando se conoce el número de grupos con anterioridad es decir a priori y los datos a clasificar son muchos además se lo utiliza cuando se quiere refinar el resultado obtenido mediante un método jerárquico.

Para implementarlo existen varias formas como lo proponen (Ruiz, Díaz, Chang, & Echeverría, 2009) pero todas ellas requieren los mismos pasos que son:

- 1) Seleccionar k centroides donde k es el número de grupos deseado
- 2) Se ordena cada observación en el grupo cuyo centroide es el más cercano
- 3) Se calculan los centroides de cada grupo
- 4) Se repiten los pasos 2) y 3) hasta que se satisfaga un criterio de parada

El método suele ser muy sensible a la solución inicial dada por lo que es conveniente utilizar una que sea buena. Una forma de construirla es mediante una clasificación obtenida por un algoritmo jerárquico

2.7.5.2 Clúster jerárquico.

Como se menciona en (IBM, 2013) los clúster jerárquicos identifican grupos relativamente homogéneos de variables basándose en las características seleccionadas, mediante un algoritmo que comienza con cada caso en un clúster diferente y combina los clústeres hasta que sólo queda uno. Es posible analizar las variables brutas o elegir de entre una variedad de transformaciones de estandarización. Las medidas de distancia o similitud se generan mediante el procedimiento Proximidades. Para seleccionar la mejor solución debemos revisar en cada etapa los estadísticos.

El análisis de conglomerados jerárquicos es una herramienta importante en el estudio de coincidencias de los modelos obtenidos respecto de una misma muestra de datos y distintas complejidades y tienen como objetivo detectar las posibles asociaciones entre los distintos modelos para que junto con otros parámetros el o los investigadores puedan discernir lo aparente de lo real como lo menciona (Pérez, 2006).

2.8 Tablas de Contingencia

Las Tablas de Contingencia son apropiadas, cuando suponemos que una variable depende de otra, estas tablas son de doble entrada donde se hace una clasificación de la muestra de acuerdo a un doble criterio de clasificación.

Implica siempre a variables cualitativas, categóricas o nominales, u ordinales con pocos valores: nominal*nominal, ordinal*ordinal; nominal*ordinal, También puede implicar a una variable nominal y otra de intervalo: nominal*intervalo(Serrano & Sánchez, 2012)

Gracias a las tablas de contingencia podemos medir la asociación entre las variables que forman la misma y una vez determinado el grado de asociación entre las dos variables, nos resta valorar si ésta es estadísticamente significativa, o lo que es lo mismo, si la asociación o relación arrojada por el estadístico elegido es debido a un error de la muestra.(Ver ANEXO 2)

Como lo menciona (Pluma, 2014) con estas tablas se puede obtener una descripción cuantitativa de las distintas cualidades bivariantes de la muestra, en forma de frecuencias y porcentajes.

2.8.1 Prueba Chi-cuadrado.

El estadístico ji-cuadrado (o chi cuadrado), que tiene distribución de probabilidad del mismo nombre, sirve para someter a prueba hipótesis referidas a distribuciones de frecuencias. En términos generales, esta prueba contrasta frecuencias observadas con las frecuencias esperadas de acuerdo con la hipótesis nula.(Quevedo Ricardi, 2011).

La finalidad de la prueba Ji Cuadrado es comprobar si existen diferencias estadísticamente significativas entre dos distribuciones de variables. Compara frecuencias observadas con frecuencias esperadas para contrastar la hipótesis de que las dos variables categóricas implicadas son independientes.(Serrano & Sánchez, 2012)

La fórmula que da el estadístico es la siguiente:

$$x^2 = \sum_i \frac{(\text{observada}_i - \text{teórica}_i)^2}{\text{teórica}_i}$$

Fórmula 1. Fórmula para calcular chi-cuadrado

Cuanto mayor sea el valor de X^2 , menos verosímil es que la hipótesis sea correcta. De la misma forma, cuanto más se aproxima a cero el valor de chi-cuadrado, más ajustadas están ambas distribuciones.

La prueba Chi-cuadrado la aplicamos para contrastar la Hipótesis Nula; H_0 = las variables x e y son independientes.

Si la significación asociada a este estadístico es menor ó igual a 0.05 rechazamos la hipótesis de independencia. La base de cálculo de la Chi-cuadrado son las diferencias entre las frecuencias observadas y esperadas.

El programa estadístico SPSS permite realizar pruebas de bondad de ajuste. Es decir, contrastar si las frecuencias observadas en cada una de las clases de una variable categórica varían de forma significativa de las frecuencias que se esperaría encontrar si la muestra hubiese sido extraída de una población con una determinada distribución de frecuencias. Un ejemplo de la prueba de Chi-cuadrado en nuestro estudio es entre el género y el lugar donde se conecta habitualmente donde comprobamos que no existe una relación entre las variables ya que el nivel de significancia es mayor al 0.05. ANEXO 3

2.8.2 Coeficiente de correlación Bivariada

La correlación bivariada es una técnica estadística que pretende determinar si dos variables tienen relación entre sí, si la relación es fuerte, moderada o débil y qué dirección tiene la relación.

La correlación bivariada está basada en la asociación lineal, es decir, que cuando los valores de una variable aumentan los valores de la otra variable pueden aumentar o disminuir proporcionalmente.

Se distinguen 2 tipos de correlaciones que son la correlación de Pearson y correlación de Spearman, están basadas en la misma información, pero utilizan fórmulas diferentes. Cuando se utiliza la curva normal se utiliza la correlación de Pearson y cuando no se la utiliza se utiliza la correlación de Spearman.

Cuando se realiza el análisis de correlación es necesario tener en cuenta el coeficiente de correlación y el nivel de significancia para determinar si existe o no una correlación entre variables.

El nivel de significancia nos determina si existe o no relación entre las variables y debe ser menos a 0.05 para poder decir que esta relación es significativa. El coeficiente de correlación oscila entre -1 y 1 mientras más se aleja del 0 existe una relación más fuerte entre las variables, esta correlación puede ser positiva o negativa dependiendo cómo afecta una variable a otra.

2.8.3 Coeficiente de correlación de Pearson

El coeficiente de correlación de Pearson nos ayuda a determinar el grado de relación lineal existente entre dos variables cuantitativas. Se lo representa por r y se lo obtiene tipificando el promedio de los productos de las puntuaciones diferenciales de cada caso (desviaciones de la media) en las dos variables correlacionadas:

$$r_{xy} = \frac{\sum x_1 y_1}{n S_x S_y}$$

Fórmula 2. Fórmula del coeficiente de Pearson

- $(x_1$ e y_1 se refieren a las puntuaciones diferenciales de cada par
- n al número de casos
- S_x y S_y a las desviaciones típicas de cada variable.

El coeficiente de correlación de Pearson por lo general toma valores entre -1 y 1: un valor de 1 indica relación lineal perfecta positiva; y el -1 indica relación lineal perfecta negativa; un valor de 0 indica relación lineal nula.

Cuando se tiene un coeficiente de correlación alto no necesariamente implica causalidad, es decir que dos variables pueden estar linealmente relacionadas sin que esto obligatoriamente me diga que una sea causa de la otra. (UCM, 2011b) Un ejemplo de relación con r de Pearson lo encontramos en el ANEXO 4, donde se analiza la correlación entre el género y las horas de conexión al día donde tiene una significancia menor a 0.05 y tiene una relación lineal negativa de -0.106.

En conclusión podemos decir que las característica principal de la correlación de Pearson es según (UV, 2011):

La correlación de una variable con ella misma siempre va a ser igual a 1.

2.8.4 Coeficiente de Determinación

Podemos definir de forma sencilla al Coeficiente de Determinación como el cuadrado del coeficiente de correlación de Pearson, y nos da la proporción de variación de la variable Y explicada por la variable X . Si la proporción es igual a 0, significa que la variable predictora no tiene capacidad predictiva de la variable a predecir (Y). Cuanto mayor sea la proporción, mejor será la predicción. Si llegara a ser igual a 1 la variable predictora explicaría TODA la variación de Y , y las predicciones NO tendrían error.

2.8.4.1 Coeficiente correlación de rangos de Kendall.

El coeficiente de correlación de rangos de Kendall, comúnmente conocida como coeficiente tau de Kendall, es una estadística que se usa para medir la asociación entre dos cantidades medidas. Una prueba de tau es una prueba de hipótesis no paramétrica para la dependencia estadística basada en el coeficiente de tau.

En concreto, se trata de una medida de correlación de rangos, es decir, la similitud de los ordenamientos de los datos cuando clasificados en orden de cada una de las cantidades. Lleva el nombre de Maurice Kendall, quien la desarrolló en 1938, cuando Gustav Fechner había propuesto una medida similar en el contexto de la serie de tiempo en 1897 (Docsetools, 2010).

2.8.4.1.1 *Tau-b- de Kendall:*

Es una extensión de Gamma pero presenta la ventaja de que alcanza los valores -1 y 1 pero sólo para tablas cuadradas cuando se presentan situaciones de total asociación.

Tau-b-de Kendall incluye los empates tomando en cuenta por separado los que surgen en la variable $X_1(Ex_1)$ y los que surgen en la variable $X_2(Ex_2)$.(UB, 2010)

$$\tau_b = \frac{C - D}{\sqrt{(C + D + Ex_1)(C + D + Ex_2)}}$$

Fórmula 3. Fórmula para calcular Tau-b

2.8.4.1.2 *Tau-c- de Kendall:*

Es una corrección de la medida anterior de modo que alcanza los valores -1 y 1 incluso en tablas no cuadradas. El problema que presenta es que tiende a subestimar la asociación real entre las variables(UB, 2010)

Este estadístico se define como

$$\tau_c = \frac{2k(C - D)}{n^2(k - 1)}$$

Fórmula 4. Fórmula para calcular Tau-c

2.8.5 **Análisis Discrimínate.**

En (F. Torrado & Berlanga, 2012) tratan al análisis discriminante como una prueba estadística adecuada para seleccionar qué variables independientes o predictivas permiten diferenciar grupos y cuántas de estas variables son necesarias para alcanzar la mejor clasificación posible. Además nos permite cuantificar su poder de discriminación en la relación de pertenencia de un sujeto u objeto a un grupo u otro. Por ello esta técnica es considerada, además de una prueba de clasificación, una prueba de dependencia. De aquí que su propósito es similar al análisis de regresión logística; la diferencia radica en que solo admite variables cuantitativas

El análisis discriminante es una técnica multivariante que se orienta principalmente a conseguir dos objetivos según (F. Torrado & Berlanga, 2012) estos son:

- a) Explicar la influencia de un conjunto de variables cuantitativas sobre una variable cualitativa.
- b) Predecir la pertenencia de los sujetos a una de las categorías de la variable criterio a partir de los valores registrados en las variables predictivas.

La variable dependiente es categórica y tiene como categorías la etiqueta de cada uno de los grupos, y las variables independientes son continuas y determinan a qué grupos pertenecen los objetos. Con estos atributos medidos se forma una combinación lineal de variables predictivas para que maximice las diferencias entre los grupos y pueda construir así un modelo predictivo para clasificar nuevos casos al grupo de pertenencia. Por tanto, el análisis discriminante será usado como lo dice (F. Torrado & Berlanga, 2012) para:

- Determinar si existen diferencias significativas entre las características de un conjunto de variables de dos o más grupos definidos a priori.
- Seleccionar las variables de predicción que contribuyen en mayor medida a explicar las diferencias entre los grupos.
- Establecer un procedimiento para clasificar a un individuo a partir de los valores de un conjunto de variables independientes.

Tras su aplicación se puede:

- Evaluar la exactitud de la clasificación en una tabla cruzada en la que se compara la pertenencia real de los individuos a un grupo con el pronóstico de pertenencia.
- Identificar qué variables independientes tienen mayor poder de discriminación y de predicción en la clasificación de sujetos.
- Obtener una ecuación denominada 'función discriminante' que exprese la combinación lineal de las variables predictivas y permita construir una regla de decisión que asigne un objeto nuevo a uno de los grupos prefijados con un cierto grado de riesgo.

2.8.6 Análisis de varianza

Según (Terrádez & Juan, 2012) el objetivo principal de muchos experimentos consiste en determinar el efecto que sobre alguna variable dependiente Y tienen distintos niveles de algún factor X variable independiente. En si para el análisis simple de la varianza consistirá en obtener muestras aleatorias e independientes del valor de Y asociado a cada uno de los distintos niveles del factor X_1, X_2, \dots, X_n . Entonces podremos determinar si los diferentes niveles del factor tienen un efecto significativo sobre el valor de la variable dependiente.

En palabras más sencillas podremos decir que vamos a utilizar el análisis de varianza para verificar si hay diferencias significativas entre medias cuando tenemos más de dos muestras en el mismo planteamiento.

Suponiendo k muestras aleatorias independientes, de tamaño n , obtenidas de una única población normal. A partir de ellas existen dos maneras independientes de estimar la varianza de la población s^2 . según (Abraira & Pérez, 1996) estas son.

- a. Una llamada varianza dentro de los grupos o varianza de error, o cuadrados medios del error, y habitualmente representada por MSE (Mean Square Error) o MSW (Mean Square Within) que se calcula como la media de las k varianzas muestrales.
- b. Otra llamada varianza entre grupos, o varianza de los tratamientos, o cuadrados medios de los tratamientos y representada por MSA o MSB (Mean Square Between). Se calcula a partir de la varianza de las medias muestrales y es también un cociente; al numerador se le llama suma de cuadrados de los tratamientos (se le representa por SSA) y al denominador $(k-1)$ grados de libertad. MSA y MSE, estiman la varianza poblacional en la hipótesis de que las k -muestras provengan de la misma población.

2.9 Regresión Logística.

Cuando se tenga una variable dependiente dicotómica que queramos predecir, o para la que queramos evaluar relación con otras variables independientes y de control, el procedimiento adecuado a realizar es una regresión logística (RL) binaria multivariante.

La Regresión Logística es el tipo de análisis multivariante más empleado en todas las Ciencias de la Vida. Las razones más poderosas según (Aguayo, 2012) son:

- a. La regresión logística nos permite introducir como variables predictores de la respuesta una mezcla de variables categóricas y cuantitativas.
- b. A partir de los coeficientes de regresión (β) de las variables independientes introducidas en el modelo se puede obtener directamente la OR de cada una de ellas.
 - Así, si la variable independiente es una variable cuantitativa, la OR que se obtiene representa la probabilidad del evento predicho que tiene un individuo con un valor x frente a la probabilidad que tiene un individuo con valor $(x-1)$.
 - Si la variable independiente es cualitativa, la RL sólo admite categóricas dicotómicas, de manera que la OR es el riesgo de los sujetos con un valor frente al riesgo de los sujetos con el otro valor para esa variable.

- c. En la RL la variable dependiente, Y, es categórica, habitualmente dicotómica (RL binaria), es decir: si o no, ocurre o no ocurre. Esto hace a este tipo de análisis el ideal para aplicar en los estudios de casos y controles, estudios en los que los casos tienen algo y los controles no.
- d. Lo que se pretende mediante la RL es expresar la probabilidad de que ocurra el evento en cuestión como función de ciertas variables, que se presumen relevantes o influyentes. Si ese hecho que queremos modelizar o predecir lo representamos por Y (la variable dependiente), y las k variables explicativas (independientes y de control) se designan por X1, X2, X3,...,Xk, la ecuación general (o función logística) es:

$$P(Y = 1) = \frac{1}{1 + \text{EXP}(-\alpha - \beta_1 X_1 - \beta_2 X_2 - \beta_3 X_3 - \dots - \beta_K X_K)}$$

Fórmula 5. Función logística

Donde α , β_1 , β_2 , β_3 ,..., β_k son los parámetros del modelo, y EXP denota la función exponencial. Esta función exponencial es una expresión simplificada que corresponde a elevar el número e a la potencia contenida dentro del paréntesis, siendo e el número o constante de Euler, o base de los logaritmos neperianos (cuyo valor aproximado a la milésima es 2,718).

Antes de hacer una regresión logística, debemos establecer claramente lo siguiente según (Aguayo, 2012):

- a. Cuáles podrían ser variables realmente predictoras (independientes) de la respuesta (dependiente). Esto lo da el conocimiento del tema y la revisión de la literatura.
- b. Cuáles podrían ser variables confundentes, que será necesario ajustar o controlar, ya que, de lo contrario, la evaluación de la relación principal ($X \rightarrow Y$) podría ser espúrea o artefactada. Esto lo da el conocimiento del tema y la revisión de la literatura, por lo que deben recogerse (e incluirse en el análisis) aquellas variables predictoras que otros estudios hayan reconocido como tales. El análisis estratificado y el análisis multivariante serán las estrategias en esta fase de análisis para corregir su efecto, procedimientos que permiten el ajuste o control.
- c. Cuáles podrían ser variables modificadoras de efecto o de interacción, que producen cambios en la relación principal evaluada ($X \rightarrow Y$) en términos de incrementarla o disminuirla. Para esto debemos amplio conocimiento del tema a tratar, aunque puede este conocimiento descubrirlo el investigador y plasmar las mismas en las conclusiones del estudio.
- d. Qué sentido tiene nuestro análisis, diferenciando dos grandes objetivos:

- a)** Predecir una determinada respuesta en base a las variables predictoras o independientes, consiguiendo una fórmula matemática que nos sirva para calcular la probabilidad de un suceso estudiado en un nuevo individuo en razón de los valores que presente de las diferentes variables incluidas en el modelo. En base a lo expuesto, se debe buscar, entre todos los posibles modelos, el más acertado, que es el que con el menor número de variables posibles genera una predicción más precisa y válida de la respuesta evaluada. Hay que tomar en cuenta que intentar desarrollar un modelo con muchas variables puede desarrollarnos un problema cuando hay pocas observaciones, ocasionando estimaciones inestables y poco precisas.
- b)** Estimar los riesgos ajustados o controlados para cada variable independiente. En este caso es importante determinar el conjunto de variables que será oportuno controlar en el análisis, incluyendo aquellas que tengan una adecuada justificación teórica. Los pasos a seguir según (Aguayo, 2012) serían:
- Valorar si hay interacción entre alguna de las variables de control y la variable independiente, con pruebas de significación estadística, dejando en el modelo los términos de interacción que sean estadísticamente significativos.
 - Valorar si hay confusión entre alguna de las variables de control y la relación principal evaluada, sin aplicar pruebas de significación estadística. En esta situación lo que debe analizarse es introducir una variable de control en el modelo de RL produce un cambio sumamente importante en la medida de asociación que estima el efecto de la exposición (X) sobre la respuesta (Y). Si esto no sucede dicha variable de control debe eliminarse del modelo, pues al dejarla puede disminuir la precisión del estudio.
 - Si al terminar el proceso existe más de un subconjunto de variables de control que ofrecen un similar grado de ajuste, se deberá tomar el que estime con mayor precisión el efecto principal evaluado ($X \rightarrow Y$) en la investigación.

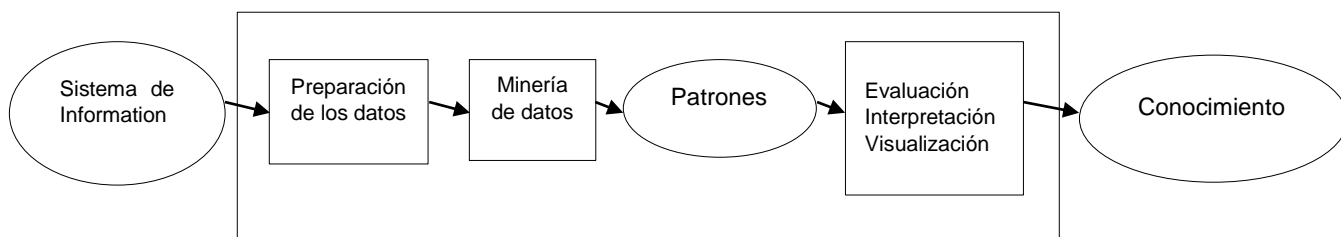
CAPITULO III
METODOLOGÍA

En este capítulo se va a describir las fases del proceso de KDD, se presenta como obtuvimos el tamaño de la muestra en la Universidad Católica Santiago de Guayaquil y que técnicas utilizamos para el levantamiento de la información para su posterior análisis.

3.1 Proceso metodológico

El proceso metodológico que se va a utilizar en este proyecto se lo detalla en la Figura 2:

Figura 2. Proceso metodológico



Fuente: (Moreno & Martínez, 2011)

a. Sistema de Información

En el presente trabajo se realizó el análisis a una porción del total de estudiantes de la Universidad Católica Santiago de Guayaquil, aplicando 500 encuestas a alumnos escogidos de forma aleatoria de varias especialidades, las encuestas que tenían información incompleta o atípicas fueron desechadas quedando un total de 393 encuestas para realizar el estudio que es superior al mínimo tamaño requerido para el análisis de la muestra según lo obtenido en el apartado 3.2.2 de este estudio.

b. Preparación de los datos

En esta etapa se determinó con que variables se va a trabajar, se utilizaron todos los atributos que se solicitan en la encuesta es decir información personal, socioeconómica y académica de los alumnos.

c. Minería de datos.

En este proyecto los procesos de minería de datos pasaron por las siguientes fases:

- Comprensión entorno estudiantil y del problema que se quiere resolver.
- Filtrado de datos
- Selección de variables.
- Extracción del conocimiento.
- Interpretación y evaluación.

d. Patrones

Al analizar los datos obtenidos se logra sacar patrones en los estudiantes sin importar la carrera que sigan más bien estos dependen de otros factores como la edad genero e ingreso económico.

e. Evaluación, Interpretación y Visualización

Se evalúan los resultados si su información es veraz para posterior interpretarlos y poder presentarlos.

f. Conocimiento

Después de analizar los resultados se logra obtener un modelo que permitiera predecir con alta fiabilidad, el rendimiento académico y los factores que afectan al mismo dentro de los estudios universitarios.

3.2 Universo y Tamaño de la muestra de Investigación

3.2.1 Universo

El universo sobre el cual se realizó la Investigación son los estudiantes de la Universidad Católica Santiago de Guayaquil cuyo número total de alumnos es 19418 al 20 septiembre del 2014.

3.2.2 Tamaño de la muestra de investigación

Sobre el valor total de alumnos de la Universidad Católica Santiago de Guayaquil vamos a calcular el tamaño de la muestra utilizando la fórmula de cálculo de tamaño de muestra para poblaciones finitas (Aguilar-Barojas, 2005) ya que se conoce el valor total de nuestra población, como desconocemos el valor exacto de la probabilidad o no que ocurra un evento vamos a suponer que estos valores son iguales es decir $p=q=0.50$. El nivel de confianza es elegido por el investigador como lo dice (Vivanco, 2005) por lo que se escogió un nivel de confianza 95% que corresponde a un índice de confianza de 1.96 y se acepta un margen de error del 5% ya que si el margen de error supera el 10 % nuestra información carecería de validez.

$$n = \frac{NZ^2pq}{E^2(N-1) + Z^2pq}$$

Fórmula 6. Fórmula para poblaciones finitas

Donde:

n= tamaño de la muestra.

N=valor de población.

Z = Nivel de confianza. (95% = 1.96)

p = Proporción proporcional de ocurrencia de un evento.

q = Proporción proporcional de no ocurrencia de un evento

E = equivale al error aceptado. 5%= 0.05

$$n = \frac{NZ^2pq}{E^2(N-1) + Z^2pq}$$

$$n = \frac{19418 * 1.96^2 * 0.5 * 0.5}{0.05^2(19418 - 1) + 1.96^2 * 0.5 * 0.5}$$

$$n = \frac{18649.05}{48.5425 + 0.9604}$$

$$n = \frac{18649.05}{49.503}$$

$$n = 376.7$$

Al aplicar la formula se obtiene 376.7 que redondeando nos da un valor de 377 encuestas que deben realizarse a los alumnos de la Universidad Católica Santiago de Guayaquil sin embargo hemos aplicado un total de 393 encuestas y sobre estas se realizara el estudio para efectos del proyecto.

3.3 Técnica de recolección de Información

Para obtener todos los datos necesarios para realizar la investigación en los alumnos de la Universidad Católica Santiago de Guayaquil se hizo uso de dos métodos de investigación, debido a que son los más conocidos, son de fácil aplicación y permiten obtener información concreta directamente de la fuente en este caso de los estudiantes y de un directivo de la universidad, estos métodos de investigación son:

- **Entrevista**

La entrevista la realizamos a la Magister María Verónica Peña Seminario quien se desempeña como Directora de Vinculación y Relaciones Internacionales de la Universidad Católica Santiago de Guayaquil, quien nos ayudó con toda la información necesaria sobre su universidad para poder realizar esta Investigación.

- **Encuesta**

Las encuestas son aplicadas a 500 alumnos de esta universidad pero después de hacer una limpieza de encuestas se las dejo en la cantidad de 393, en estas encuestas se trata de obtener información referente a:

- Aspecto sociodemográfico de los alumnos con información como edad, género y nivel de ingresos.
- Información sobre los perfiles de conexión de los alumnos, a través de variables lugar de conexión, nivel de conocimiento, días de conexión a la semana, nivel de conocimiento en el manejo de Internet, cuantas horas ser conecta al internet diariamente, y años de experiencia en la conexión a internet.

- Información sobre las actividades de índole académica que realiza el estudiante en línea.
- Información sobre las actividades de entretenimiento que realiza el estudiante en línea.
- Información sobre si el estudiante cuenta con cuentas en redes sociales y páginas tanto de entretenimiento como de índole académica, y la cantidad de seguidores o contactos que tiene en las mismas.
- Información sobre el nivel de uso de dispositivos tecnológicos por parte de los estudiantes.
- Información sobre la percepción del estudiante en la utilidad e importancia del uso de Internet.
- Información sobre la percepción el uso de los recursos educativos en línea por parte de los profesores en las actividades académicas.

El formato de esta encuesta y las preguntas ahí realizadas las podemos encontrar en el (Ver ANEXO 1)

3.4 Obtención de los resultados

Como lo explica (Espíndola & León, 2002) primero previa la aplicación de una de las técnicas de minería de datos, tenemos que realizar algunas tareas de pre-procesado, que nos permitan transformar los datos originales a una forma más apropiada para ser usada por el algoritmo en particular. Para este proyecto las tareas realizadas han sido la integración, la limpieza, la transformación y la discretización de datos. La integración de los datos ha consistido en la agrupación de toda la información obtenida por medio de las encuestas a cada estudiante en una base de datos para su posterior análisis.

En la etapa de limpieza, se elimina a las encuestas que no tienen completa al 100% toda la información o tienen datos atípicos. En la transformación de datos se cambian varios nombres de atributos e instancias que contenían la letras en mayúsculas o minúsculas para estandarizar, se agrupa a los estudiantes en la pregunta 12 de la encuesta de acuerdo al número de seguidores que tienen en sus diferentes cuentas de Facebook, twitter y LinkedIn, además se agrupo a los estudiantes quienes aprobaron y quienes reprobaron. En la etapa de discretización, el uso de las tecnologías en actividades académicas se transforma por medio del análisis clúster en estudiantes de uso Bajo, bueno y regular, en el uso de internet en actividades de entretenimiento se las clasifíco en decreciente, creciente y constante y en el rendimiento académico en estudiantes aplicados y descuidados (Vera, Morales, & Soto, 2012). Se creó un fichero con formato .sav de SPSS que es el software que utilizamos para realizar la minería de datos.

CAPITULO IV
RESULTADOS DE LA INVESTIGACIÓN

En este capítulo se va a representar los datos que se obtiene de la investigación realizada a nuestro universo de investigación se va a representar en gráficas para su mejor visualización y posterior interpretación.

4.1 Resultados

Una vez que se realizadas las primeras fases del KDD es decir la integración y recopilación de datos es decir donde una vez realizadas las encuestas manualmente a los estudiantes de la Universidad Católica Santiago de Guayaquil, se utilizó la herramienta en línea SurveyMonkey para poder tener la información dentro de una base de datos, posteriormente se realiza la segunda fase del KDD que es la selección, limpieza y transformación de datos donde eliminamos estudiantes con información incompleta o datos atípicos que pueden afectar a los resultados posteriores para la comprobación de las hipótesis. Una vez realizada la limpieza de datos realizamos la transformación de los mismos para poder trabajar con ellos de una manera adecuada y posteriormente seleccionar los datos más importantes y que influyan para el cumplimiento de nuestras hipótesis.

4.1.1 Datos Generales de los estudiantes.

Al realizar la investigación en la Universidad Católica Santiago de Guayaquil se evidencia que la muestra ha sido realizada en alumnos de la mayor cantidad de carreras ofertadas por esta universidad, siendo los alumnos de mayor frecuencia los pertenecientes a la carrera de medicina con 11.2% del total de la muestra como se observa en la figura Nro. 3.



Figura 3. Carreras Universitarias

Fuente: Elaboración propia

En la tabla Nro. 2 se representa la información obtenida de la muestra referente a la edad donde se evidencia que la población universitaria es relativamente joven al tener una media de 20.32 y una desviación de 2.33.

Tabla 2. Datos de la edad

Estadísticos		
¿Cuál es su edad?		
N	Válidos	393
Edad mínima		17
Edad máxima		37
Media		20.32
Mediana		20.00
Moda		20
Desv. típ.		2.337
Varianza		5.463

Fuente: Elaboración propia

De la población estudiada el 62% pertenece a las mujeres y 38 % a los hombres.

El 73% de los estudiantes tienen en su familia ingresos altos, considerando para esto los estudiantes cuyas familias tiene ingresos superiores a 600 dólares, el 27 % restante tienen unos ingresos menores a 600 dólares como se observa en la Figura Nro.4.

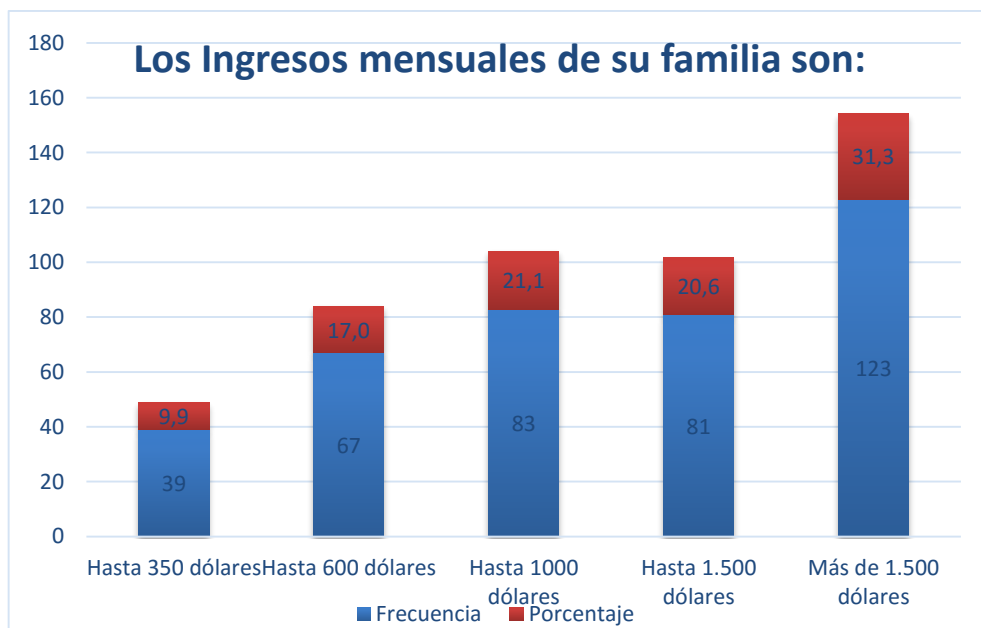


Figura 4. Distribución de estudiantes por los Ingresos mensuales

Fuente: Elaboración propia

4.1.2 Hábitos de conexión de la población estudiada

La mayor cantidad de los estudiantes se conectan desde la casa formando el 56.7% del total la muestra al igual que desde una red móvil con un 23.9%, de la universidad un 17.6% del total, mientras que desde donde menos se conectan formando el 1.8% es desde el trabajo y desde un cyber café.

Considerando que aquellos estudiantes que indican tener un conocimiento igual o superior a siete de una escala de 1 al 10 (Figura 5), tenemos un 88.8% de la muestra con un conocimiento alto en el manejo de internet, este nivel de conocimiento viene dado por la cantidad de días y horas que pasan conectados, se encuentra que el 85% de la muestra se conecta 7 días a la semana, con un promedio de conexión diario de 8 horas.

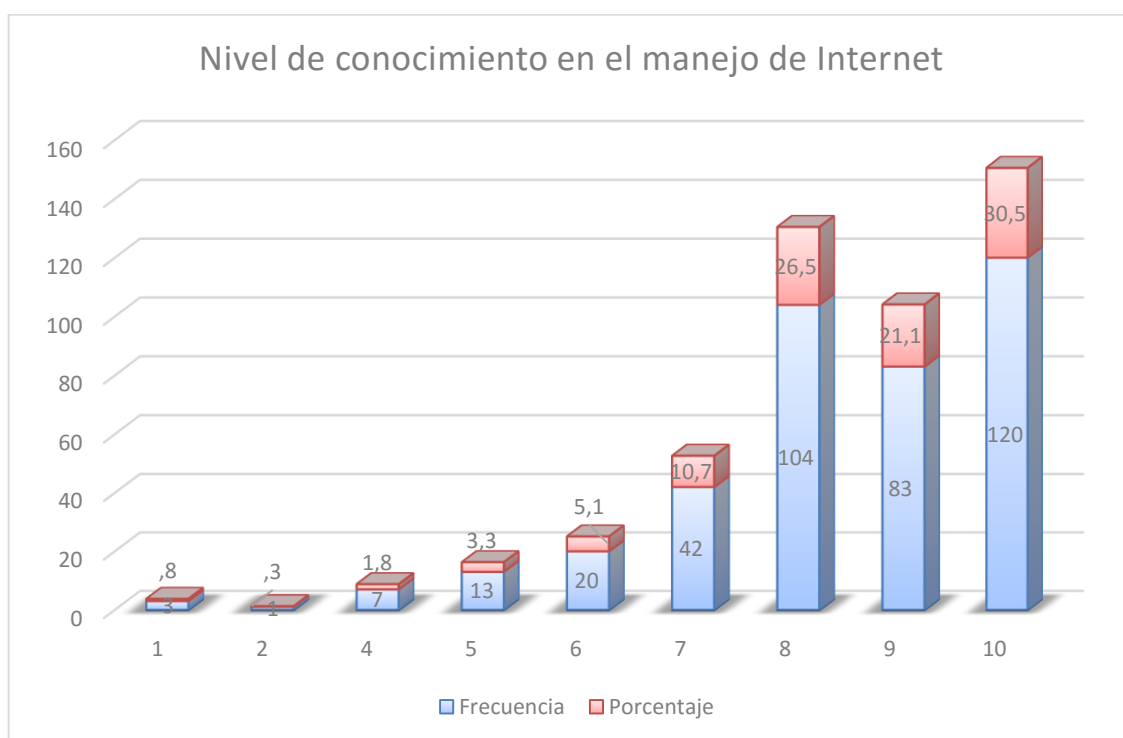


Figura 5. Distribución de estudiantes por el nivel de conocimiento en el manejo de internet
Fuente: Elaboración propia

4.1.3 Uso de Internet en Actividades Académicas

Aunque el 85% de los estudiantes se conecten siete días a la semana el ingreso a la plataforma virtual es bajo como se observa en la figura Nro. 6 de donde obtenemos un promedio de 2 ingresos semanales a la plataforma virtual, el poco ingreso a la plataforma virtual deriva en la baja participación en foros virtuales pues el 47% de la muestra no participa en los mismos y el 53% restante tiene una media de 1 participación mensual.

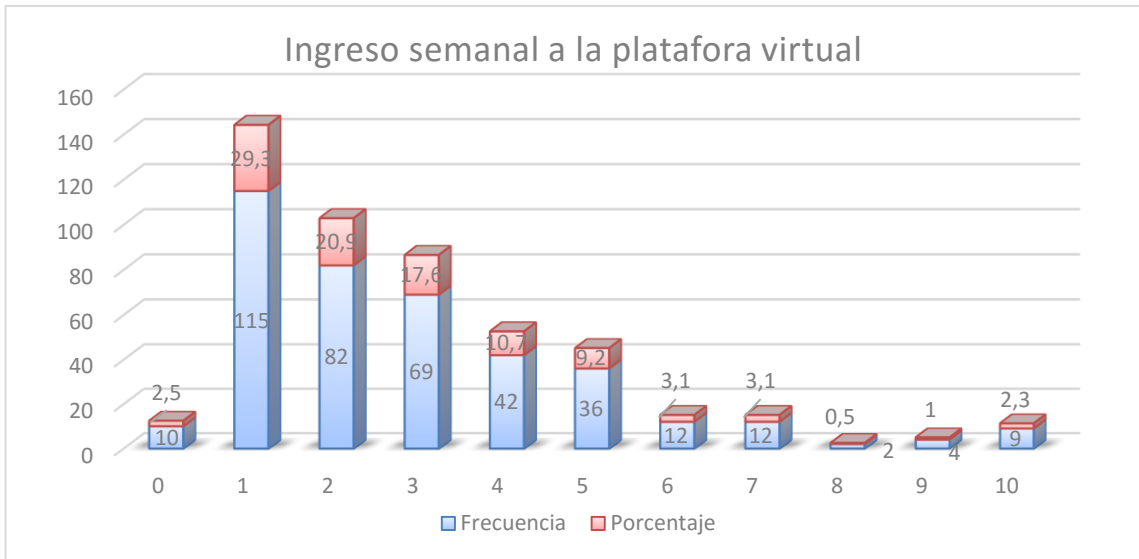


Figura 6. Ingreso a la plataforma virtual

Fuente: Elaboración propia

En la figura Nro. 7 se observa que las consultas realizadas a los profesores son muy bajas y de donde se obtiene una media de 4 consultas mensuales a los profesores, siendo estas menores al promedio de consultas realizadas a los compañeros que es de 16 consultas, que nos muestra que no hay confianza suficiente entre docentes y alumnos y más bien hay una interacción mayor entre compañeros.



Figura 7. Consultas al profesor

Fuente: Elaboración propia

La descarga de recursos educativos al igual que el número de videos académicos vistos en internet es baja considerando el promedio de conexión diaria pues solo se tiene una media de 3 descargas de recurso educativos y una media de 4 videos educativos vistos mensualmente.

En la figura Nro. 8 se representan las horas de chat sobre temas académicos, de aquí obtenemos una media de 5 horas de chat mensuales sobre temas académicos, de donde resulta lógica la baja cantidad de post o tweets académicos y por ende la poca interacción entre compañeros y docentes sobre temas académicos.

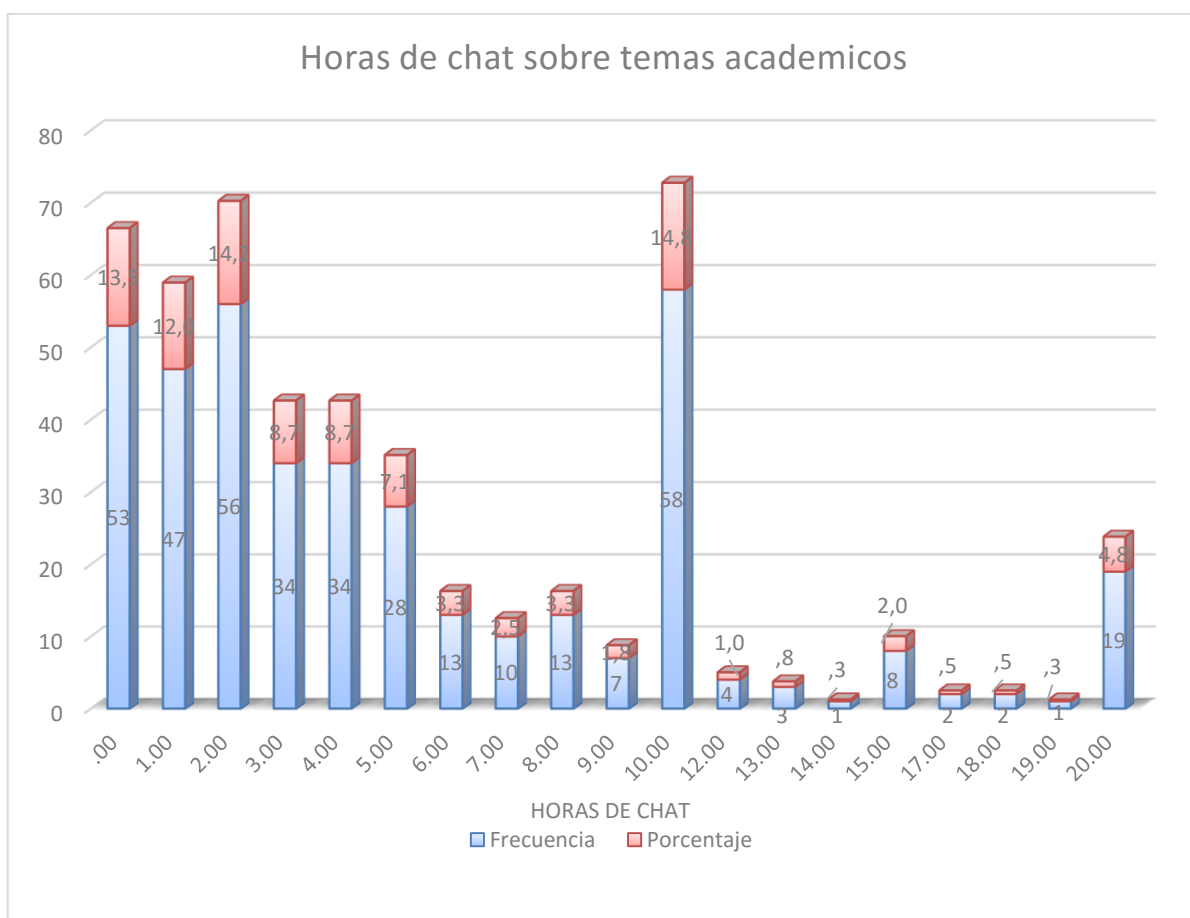


Figura 8. Chat temas académicos

Fuente: Elaboración propia

Considerando las horas de conexión mensuales que son 224 horas, es interesante que de estas solo en promedio 12 horas sean usadas para la búsqueda de información académica al mes, de estas doce horas un promedio de 4 horas son utilizadas en la búsqueda de información académica en la biblioteca de la universidad lo que nos demuestra que los alumnos tienen otras fuentes de búsqueda indistintas a las brindadas en la universidad.

El 89.1% de los estudiantes expresa que internet le permite realizar los trabajos más rápido y con menor esfuerzo además indican que internet les facilita los procesos de aprendizaje por lo que los estudiantes han prescindido de la biblioteca tradicional en un 67.9% del total de la

muestra, pero la confianza en la información obtenida de la misma baja con respecto al porcentaje anterior pues solo el 70.5% confía en la información brindada por internet como se observa en la figura 9.

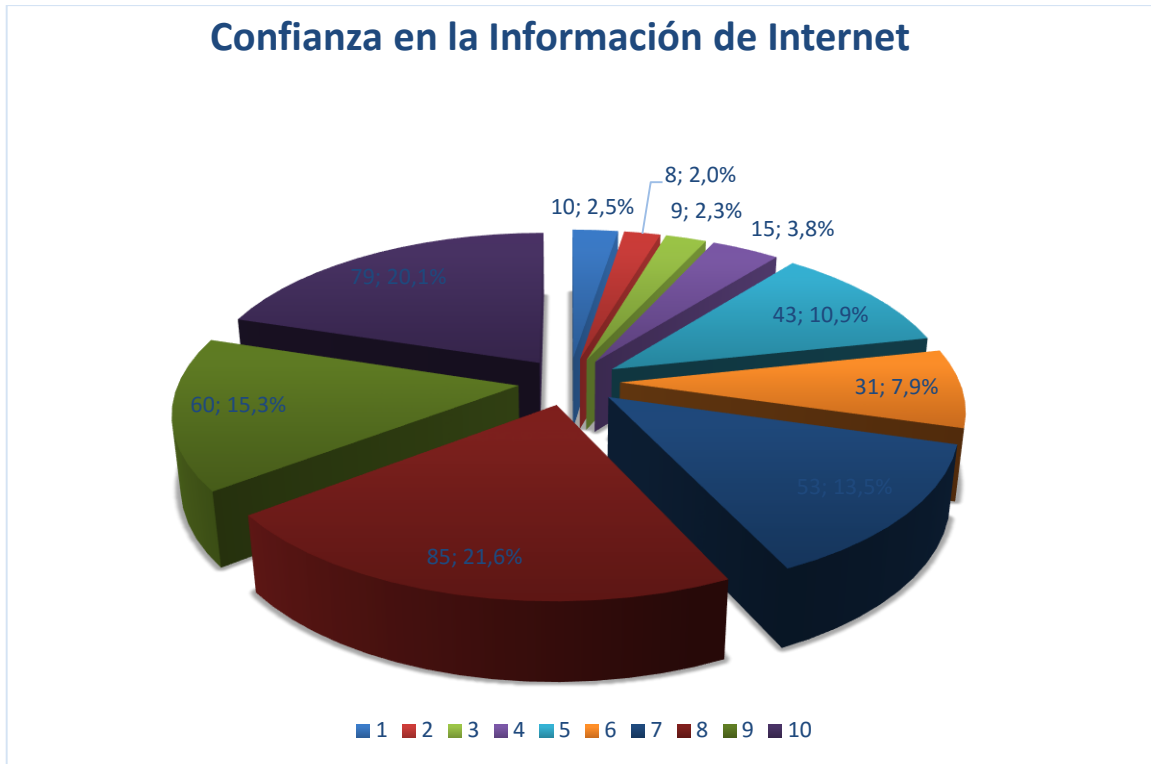


Figura 9. Confianza en la Información de Internet
Fuente: Elaboración propia

La gran mayoría de estudiantes es decir un 74% de ellos cree que internet le permite mejorar sus calificaciones pero tenemos una gran cantidad de estudiantes que realizan sus trabajos copiados desde internet como se observa en la figura 10, lo que tiene aspectos negativos ya que en su mayoría no realizan una correcta discriminación de la información encontrada en internet sino más bien realizan un proceso de copiar y pegar simplemente para cumplir con el trabajo.

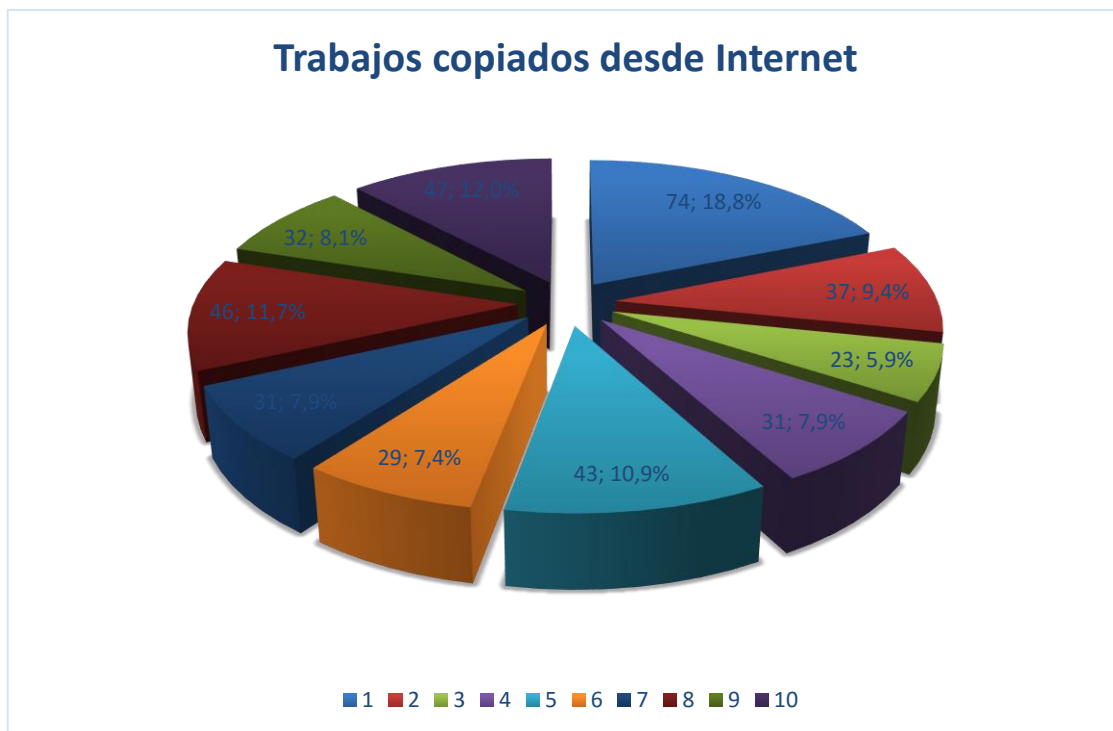


Figura 10. Trabajos copiados desde Internet
 Fuente: Elaboración propia

4.1.4 Uso del Internet en actividades de Entretenimiento

Es importante destacar que más del 90% de la población universitaria estudiada posee cuentas tanto en Facebook como twitter, por lo que su interacción en estas redes sociales va a ser alta.

En la figura Nro. 11 se representa la distribución de estudiantes por horas de chat por diversión a la semana de donde obtenemos una media de 11 horas semanales que es superior al promedio de chat por temas académicos que es 5, donde podemos darnos cuenta que los estudiantes utilizan más el chat para la interacción social que para temas académicos.

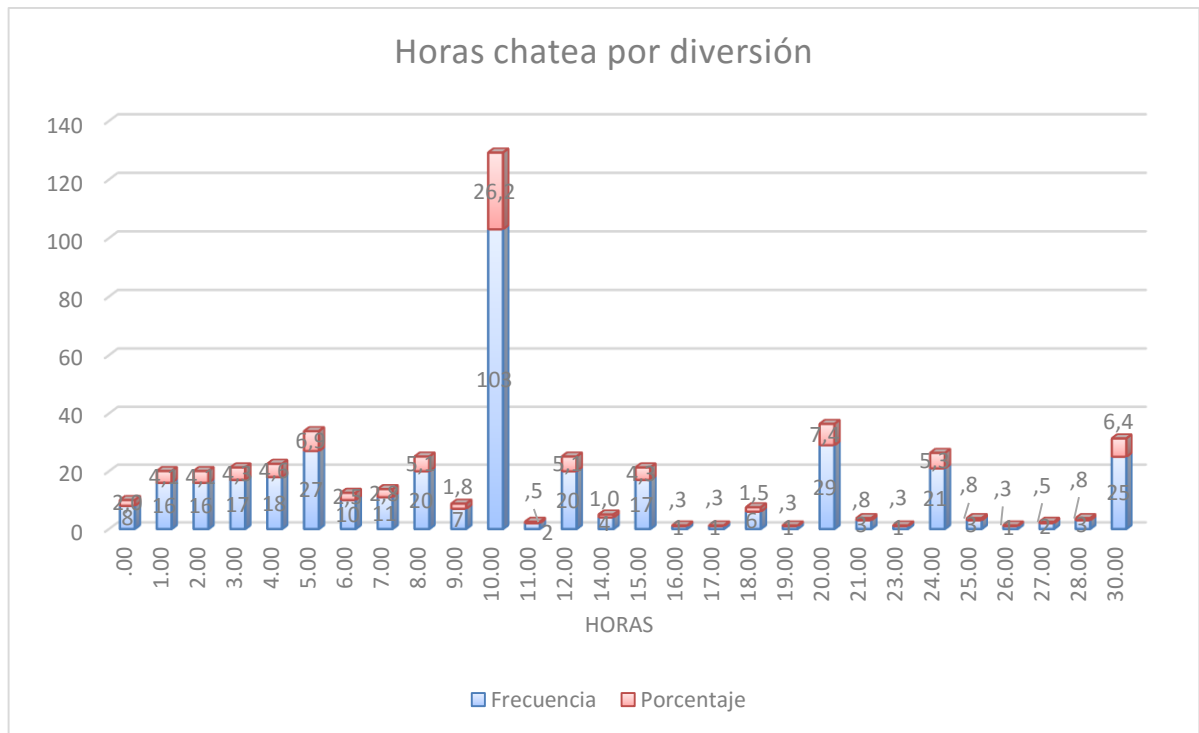


Figura 11. Chat por diversión

Fuente: Elaboración propia

El uso de redes sociales en la población universitaria tiene una media de 13 horas semanales que va concatenado con las horas que chatea por diversión

En la población universitaria estudiada encontramos que los juegos en línea tiene una media de 3 horas semanales que es bajo comparando las horas que se conecta al día, esto se da también pues el 62% de la muestra pertenece al género femenino y ellas no son afines a los juegos en línea

La descarga de contenidos de la web como música, videos y programas y el ver videos por entretenimiento tienen promedio semanales bajos, la descargas de contenidos llega a una media de 4 y el ver videos por entretenimiento una media de 8 videos que no se contrasta con el hecho de que el 70 % de los estudiantes posean cuenta en YouTube.

4.1.5 Uso de Dispositivos Tecnológicos

Considerando que los valores superiores a siete entre una escala de 1 a 10 son altos, tenemos que el 87.3% de la población universitaria tiene un nivel de uso alto de un Smartphone, así como el 84.6% tiene un nivel de uso alto de teléfono con acceso a internet. Lo que no sorprende es el bajo uso de dispositivos móviles sin acceso a internet pues supera el 60.5% del total de la población universitaria, figura 12.

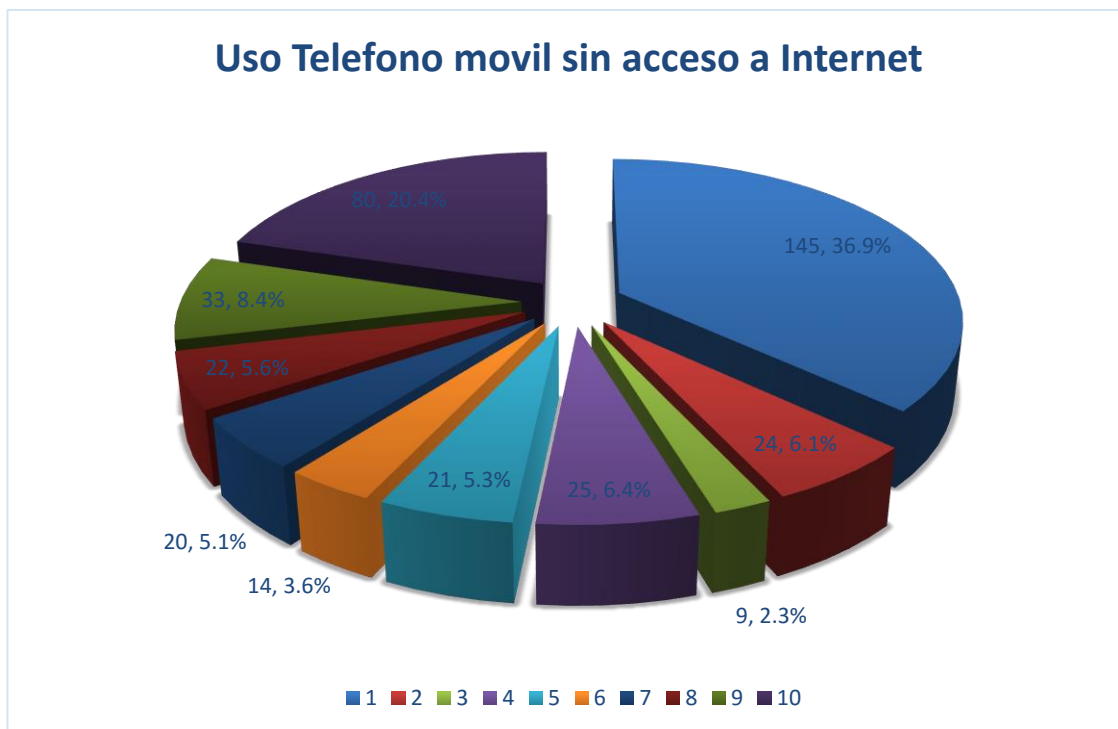


Figura 12. Uso teléfono móvil sin acceso a Internet
 Fuente: Elaboración propia

Considerando que la muestra posee ingresos económicos altos y por ende las facilidades para poseer un computador portátil, que es una herramienta necesaria para el desarrollo de actividades académicas, encontramos que el 73% de la muestra posee un nivel de uso alto de la misma

El 60% de los estudiantes investigados tienen un nivel alto de uso de una Tablet, el uso de la cámara digital, iPod, mp3 es menor debido a que los Smartphone ya cuentan con cámaras inteligentes y además se puede reproducir tanto música como videos en los mismos.

4.1.6 Interacción entre docentes y estudiantes.

Aunque el 88% de los estudiantes indican que sus profesores ingresan a la plataforma virtual y que un 80% los mismos suban materiales digitales a la plataforma virtual y otro 75 % recomienda recursos de la biblioteca virtual, según las los datos obtenidos en esta investigación, no concuerda con el poco ingreso e interacción de los estudiantes en la esta plataforma así como el poco ingreso registrado en la bibliotecas virtuales.

El 66% de los estudiantes afirma que sus profesores si plantean cuestionarios y evaluaciones en la plataforma virtual así como también les plantean foros pero su poco ingreso a las mismas hacen suponer que estas evaluaciones y foros no tienen una ponderación elevada para que la gran mayoría de estudiantes no las rinda o no participe de estos foros.

A pesar de que más del 65% de docentes de esta universidad cuenta con correos electrónicos y redes sociales la interacción entre estudiantes y profesores es muy baja.

4.1.7 Rendimiento Académico

De los datos obtenidos en la figura Nro. 13 obtenemos que la media en materias matriculadas en el semestre anterior es de siete.

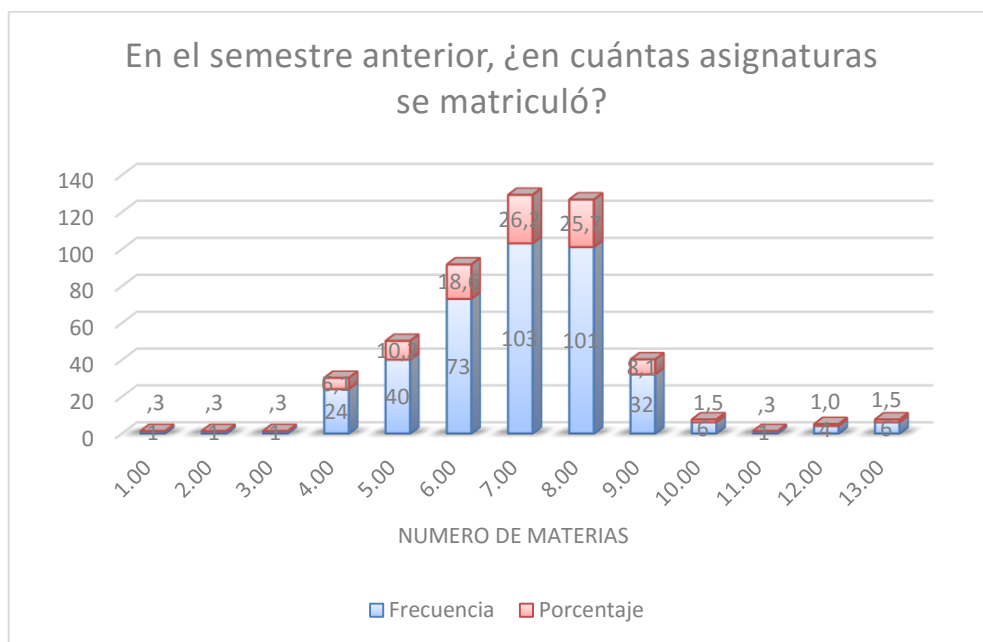


Figura 13. Materias que se matriculo

Fuente: Elaboración propia

Así también obtuvimos de la figura 14 la media de materias aprobadas en el semestre anterior es de siete materias aprobadas en el semestre anterior.

La cantidad de alumnos que han reprobado materias en esta institución educativa es baja ya que más del 80% de los estudiantes no han perdido ni una materia y esto lo obtenemos restando el número de materias matriculadas menos el número de materias reprobadas de cada estudiante.

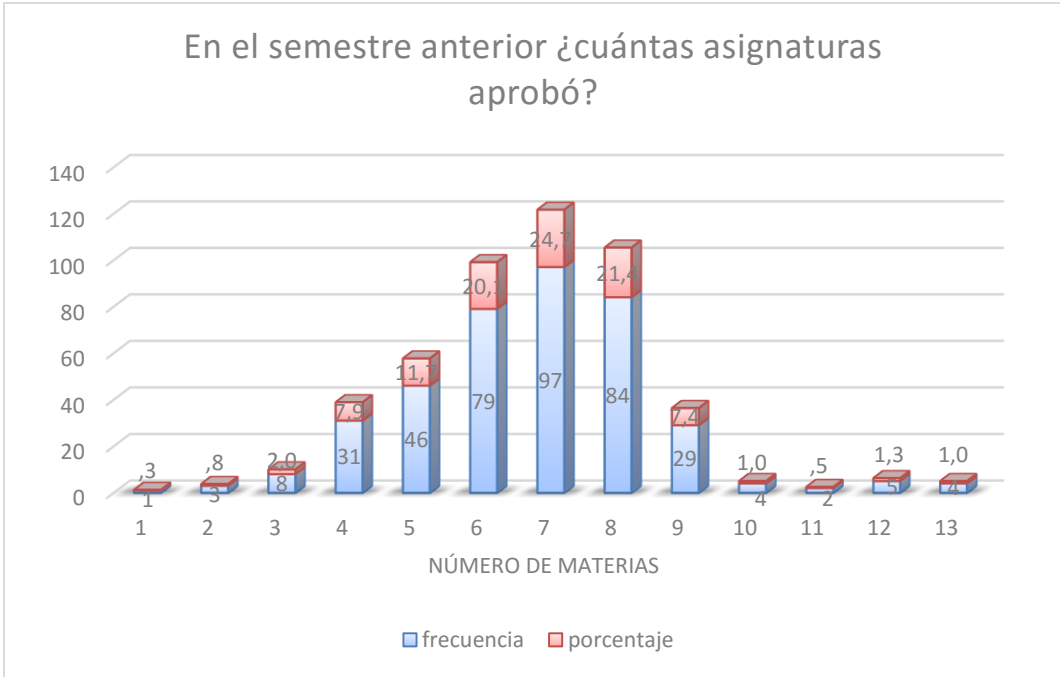


Figura 14. Materias que aprobó

Fuente: Elaboración propia

CAPITULO V
ANÁLISIS DE RESULTADOS Y COMPROBACIÓN DE HIPÓTESIS

En este capítulo se va a presentar los resultados obtenidos en la fase de minería de datos en donde se realiza el análisis de correlaciones entre las variables obtenidas de la investigación, además se describe los resultados obtenidos al aplicar análisis factorial y análisis clúster, para comprobar las hipótesis se realizar una regresión logística multinomial y binomial.

5.1 Relaciones entre variables

El excesivo uso del Internet, definido según (Ferran, 2009) por la cantidad de horas que una persona permanece navegando en la red, va afectar directamente en la vida social y estudiantil de los adolescentes, por lo que es importante analizar qué relación tienen la edad, género e ingresos económicos mensuales sobre las variables obtenidas en nuestra investigación pues el encontrar alguna influencia de una variable sobre otra se va a poder ir comprobando las hipótesis planteadas. Se ha escogido el género porque como menciona (Sánchez Galvis, 2010) es necesario realizar un análisis del acceso y uso diferencial de las TIC entre hombre y mujeres, pues las tecnologías son un potencial motor de desarrollo, superación e inclusión y pueden influir positiva o negativamente en la brecha de género existente. La edad también influye significativamente en el uso de las tecnologías pues el uso de las mismas de un adolescente va a ser diferente al uso que le dé una persona mayor, así como el nivel de ingresos económicos permite adquirir recursos tecnológicos actuales para el desarrollo de las actividades tanto académicas como de entretenimiento.

Para determinar que variables influyen sobre otras se aplicó análisis de correlaciones¹, si son variables numéricas se va aplicar análisis de correlaciones de Pearson, si son dos variables categóricas se va aplicar chi-cuadrado, si se va a relacionar una variable categórica con una numérica tenemos dos puntos a tomar en cuenta si es una relación cuadrada se aplicara Tau-b de Kendall y si no lo es se aplica Tau-c de Kendall. Para todos estos análisis de correlaciones vamos a tomar en cuenta un nivel de significancia igual o menor a 0.05, y se tomara las variables que tengan mayor coeficiente de correlación que este más alejado de 0.

5.11.1.1 Análisis de la incidencia de la edad sobre resultados obtenidos.

Al aplicar análisis de correlaciones se determina que la edad incide sobre algunas variables, las que la igual que la edad fueron variables numéricas por lo que se tomara los datos obtenido por R de Pearson.

Como se menciona en (Albero, 2002) las investigaciones realizadas en los últimos años enfocadas ende las nuevas tecnologías de información y comunicación por parte dela población más joven, lo que expone una preocupación creciente por definir cuál es la

¹ Sección 2.8.2 Análisis de correlación Bivariada.

influencia que ejerce el uso de estos medios en todos los procesos de formación de las poblaciones jóvenes como lo menciona (Buckingham & Sefton-Green., 1999).

Se encuentra una relación significativa entre la edad y las horas que se conecta al día, como se observa en la tabla 4 tiene un nivel de significancia equivalente 0.035 y un coeficiente de correlación negativa de -0.106, el coeficiente de determinación es de 1.124 %. Esto no se aleja de la realidad ya que como afirma (Albero, 2002) los estudiantes y adolescentes en general han incorporado el uso de internet al conjunto de sus actividades cotidianas. La importancia de Internet como una fuente de acceso infinita al conocimiento no necesariamente es lo que ha llevado a las poblaciones jóvenes a conectarse de manera permanente sino más bien han encontrado en Internet un herramienta más accesible para enfocar sus intereses en común con los grupos de su misma edad. Esto se contrasta con los resultados que hemos encontrado en nuestra investigación en la sección del uso de internet en actividades académicas², donde se evidencia las pocas horas de uso del internet para estas actividades.

En la investigación se encuentra una relación significativa entre la edad y las horas que busca información académica en internet cada mes, con un nivel de significancia equivalente a 0.029 y un coeficiente de correlación de 0.110 como se ve en la tabla Nro. 3. El coeficiente de determinación es de 1.21 %, a pesar de esta relación positiva entre estas dos variables el promedio mensual de búsqueda de información académica es de 16 consultas que es baja tomando en cuenta que es un promedio mensual, esto se da porque la edad promedio de la población estudiada es de 20 años es decir son relativamente jóvenes, si tomamos analizamos los estudiantes con mayor edad dentro de esta investigación podemos darnos cuenta que a mayor edad mayor número de horas de búsqueda información por ejemplo tomando en cuenta a estudiantes que tiene 24 años el promedio mensual de búsqueda sube a 18 horas.

Tabla 3. Variables que tienen correlación con la edad

	R de Pearson	Sig. Aproximada
Horas de Conexión	-0.106	0.035
Años de Experiencia	0.278	0.000
Horas de búsqueda de Información académica	0.110	0.029
Seguidores en Twitter	-0.123	0.015

Fuente: Elaboración propia.

² Literal 4.4 Uso del internet en actividades académicas.

Se encuentra una relación significativa entre la edad y los años de experiencia, tiene un nivel de significancia equivalente a 0.000 y un coeficiente de correlación de 0.278 como se ve en la tabla Nro. 3. El coeficiente de determinación es de 7.728 %, es evidente y lógico esta relación pues a mayor edad, y al estar envuelto en el actual entorno educativo especialmente, va a tener mayor experiencia en el uso de internet.

Aunque las redes sociales están presentes en la vida cotidiana, en nuestro caso de estudio solo se encuentra una relación significativa entre la edad y seguidores en twitter, con un nivel de significancia equivalente a 0.015 y un coeficiente de correlación negativo de -0.123 como se ve en la tabla Nro. 3. El coeficiente de determinación es de 1.513 %. Aunque muchos autores como (Ebner, Lienhardt, Rohs, Meyer, & I., 2010) reconocen que twitter es útil en el ámbito universitario al facilitar la difusión de información y fomentar el debate entre estudiantes interesados en un tema específico, se hace interesante la relación inversa entre estas dos variables y se comprueban con los resultados obtenidos en el levantamiento de información de nuestra en donde se demuestra que existe una media de 3 tweets sobre temas académicos al mes al tener una edad promedio de 20 años.

5.11.1.2 Análisis de la incidencia del género sobre Variables.

El género incide sobre algunas variables después de aplicar un análisis de correlaciones, que es parte de la minería de datos, a la variable género se la va a tomar como una variable categórica por lo que se va a utilizar Chi- cuadrado cuando se relaciones con otra variable categórica, Tau-b y Tau-c de Kendall para las relaciones con variables numéricas.

Tabla 4. Variables que tienen correlación con el género.

	Chi Cuadrado	Tau-c de Kendall	Sig. Aproximada
Horas de conexión diaria		0.112	0.043
Horas de juego en línea		-0.158	0.003
Horas de descarga de contenidos		-0.111	0.046
Uso del teléfono móvil		0.122	0.017
Materias aprobadas		0.120	0.029

Fuente: Elaboración propia.

Trabajos como el de (Castaño, Fernández, & Vásquez, 2008) insisten, que los hombres utilizan Internet más que las mujeres, los hombres por lo general son más propensos al uso de internet para juegos, actividades de consulta, pero no necesariamente relacionada con la educación sino más bien con el deporte, o a la descargar de varios contenidos de la web,

mientras que las mujeres por el contrario sienten más atracción a la ejecución de actividades más funcionales, relacionadas con aspectos como la salud, la educación, formación y la relación familiar, estas diferencias se basan en general a la múltiple elección de actividades por parte del género femenino en comparación son los hombres.

Se encuentra una relación significativa entre el género y cuantas horas se conecta cada día al ser una relación de una variable categórica con una numérica Tau-c de Kendall con un valor de 0.112 y un nivel de significancia de 0.043, como se observa en la tabla Nro. 4. Tomando en cuenta que la mayor cantidad de la muestra pertenece al género femenino.

Se encuentra una relación significativa entre el género y cuantas horas a la semana utiliza juegos en línea con un valor de Tau-c de Kendall equivalente a -0.158 y un nivel de significancia de 0.003 como se ve en la tabla Nro. 4, es decir que los hombres van a jugar más que las mujeres en línea, lo que se contrasta con lo expuesto por (Castaño et al., 2008).

Al ser la mayor parte de la muestra mujeres es lógico encontrar una relación significativa negativa entre el género y cuantas horas a la semana descarga contenidos como música, videos y programas con un valor de Tau-c de Kendall equivalente a -0.111 y un nivel de significancia de 0.046 como se observa en la tabla Nro. 4, puesto que por diferentes estudios realizados como los de (Gargallo, Suárez, & Belloch, 2003) se ha demostrado que los hombres tienden a realizar más descargas de contenidos en internet que las mujeres.

Se encuentra una relación significativa entre el género y el nivel de uso de un teléfono móvil con un valor de Tau-c de Kendall igual a 0.122 y un nivel de significancia de 0.017 como se ve en la tabla Nro. 4, esta relación se da pues el uso móvil en la actualidad según (Carbonell & Beranuy, 2007) el teléfono móvil es un símbolo de estatus dentro de los estudiantes universitarios y la universidad objeto de nuestra investigación es una universidad de una población con un estatus económico alto por ende el celular y su uso es parte del diario vivir para sus diversas actividades tanto estudiantiles como cotidianas.

Como expone en (Caso & Hernández, 2007) el género femenino responde con más eficacia a las exigencias del entorno estudiantil y tienden a desenvolverse de una forma más eficiente en ambientes estructurados, son las mujeres quienes más se interesan en el estudio, se organizan de mejor manera todas sus actividades académicas, tienen un mayor sentido de ayuda al momento de compartir estrategias y técnicas que favorecen su procesos de aprendizaje y la comprensión lectora, además las mujeres presentan mayor habilidad para fijar metas personales y profesionales por lo que nuestro caso de estudio no se aparta de la realidad ya que se encuentra una relación significativa entre el género y materias aprobadas

con un valor de Tau-c de Kendall igual a 0.120 y un nivel de significancia de 0.029 como se ve en la tabla Nro. 4.

5.11.1.3 Análisis de la incidencia de los Ingresos mensuales sobre Variables.

Como en los análisis anteriores se aplicó análisis de correlaciones y se encuentra que los Ingresos Mensuales inciden sobre algunas variables, para lo cual se toma en cuenta a los ingresos mensuales como una variable numérica por lo que para el análisis de correlaciones con las diferentes variables se tomó en cuenta R de Pearson, Tau -b y Tau-c de Kendall.

La variable ingresos mensuales familiares es la que más determina el comportamiento de los estudiantes en la investigación que se está desarrollando.

Se encuentra una relación significativa entre los ingresos mensuales y el lugar donde se conecta habitualmente a Internet con un valor de Tau-b de Kendall igual a 0.090 y un nivel de significancia de 0.037 como se observa en la tabla Nro. 5. El 73% de la muestra cuenta con un entorno familiar con ingresos económicos altos, por lo que cuentan con conexión a internet tanto en sus hogares como en una red móvil, aunque la conexión desde la universidad también es alta pues es nuestra investigación más nos interesa los valores anteriores ya que al realizar la investigación en una Universidad de primer nivel es lógico que cuente con una red inalámbrica propia como la mayoría de universidades a nivel nacional.

Se encuentra una relación significativa entre los Ingresos mensuales y cuantos días a la semana se conecta a Internet con un valor de R de Pearson de 0.274, un nivel de significancia de 0.000, como se observa en la tabla Nro. 5, el valor del coeficiente de determinación es de 7.508 %, al contar con una conexión tanto en sus hogares como universidades y red móvil es lógico que el 85% de los estudiantes se conecten los siete días a la semana.

Se encuentra una relación significativa entre Ingresos mensuales y el nivel de conocimiento para el manejo de Internet con un valor de R de Pearson igual a 0.144, el nivel de significancia es de 0.004, como se observa en la tabla Nro. 5, el valor del coeficiente de determinación es 2.074%, al pertenecer a entornos familiares con ingresos económicos altos la adquisición de nuevas tecnologías es parte de su vivir por lo que pueden conectarse varios días a la semana y por ende adquieren mejores destrezas en el manejo de internet y de las tecnologías, por esto también se encuentra una relación significativa entre Ingresos mensuales y cuantas horas se conecta cada día con un valor de R de Pearson igual a 0.130, el nivel de significancia es de 0.010, como se observa en la tabla Nro. 5, el valor del coeficiente de determinación es 1.69%.

Tabla 5. Variables que tienen correlación con los Ingresos mensuales

	R de Pearson	Tau b de Kendall	Sig. Aproximada
Lugar de conexión		0.090	0.037
Días de conexión	0.274		0.000
Nivel de conocimiento manejo de Internet	0.144		0.004
Horas de conexión	0.130		0.010
Años de Conexión	0.254		0.000
Horas de chat por diversión	0.146		0.004
Descarga contenidos	0.144		0.004
Uso Smartphone	0.267		0.000
Uso teléfono móvil con acceso a internet	0.144		0.000
Uso computador portátil	0.111		0.027
Uso de Tablet	0.172		0.001
Cuántas asignaturas se matriculó	-0.192		0.000
Cuántas asignaturas aprobó	-0.191		0.000

Fuente: Elaboración propia.

Se encuentra una relación significativa entre Ingresos mensuales y cuantas horas a la semana chatea por diversión con un valor de R de Pearson igual a 0.146, el nivel de significancia es de 0.004, como se observa en la tabla Nro. 5, el valor del coeficiente de determinación es 2.132%, esta relación concuerda con los resultados obtenidos en esta investigación ya que el promedio de horas de chat semanal es alto, además encontramos una relación significativa entre los ingresos mensuales y la descarga de contenidos desde la web como se observa en la tabla Nro. 5, pero la descartamos pues en los datos obtenidos el promedio de descarga es muy bajo.

En esta universidad los alumnos pertenecen a un entorno familiar con ingresos económicos altos, por lo que entre ellos el estar actualizado con los últimos dispositivos tecnológicos es parte fundamental de su diario vivir, por lo que se encontraron varias relaciones significativas entre los ingresos mensuales y el uso de varios dispositivos tecnológicos.

Como se observa en la tabla Nro. 5, se encuentra una relación significativa entre Ingresos mensuales y el nivel de uso tanto de un Smartphone con cámara fotográfica y acceso a Internet como así también una relación con un teléfono móvil con acceso a internet con unos coeficientes de determinación respectivamente de 7.129% y 4%, esta relación concuerda con los datos obtenidos de nuestra investigación ya que aproximadamente el 86% del total de la muestra tienen un nivel de uso alto de los mismos.

Al encontrar un alto número de estudiantes con un promedio de uso alto de computadoras portátiles la relación entre los ingresos mensuales y nivel de uso de un computador portátil es predecible, se lo confirma con el valor de R de Pearson igual a 0.111, un nivel de significancia es de 0.027, y un coeficiente de determinación es 1.232% como se observa en la tabla nro. 5.

Considerando que las Tablets son dispositivos con los que al igual que con un Smartphone o una Laptop podemos acceder a navegación web, e-mail, lectura y edición de documentos simples, es decir brindan al estudiante varias facilidades tanto para actividades académicas como de entretenimiento reforzando su autonomía personal y creatividad, y su utilización es alta entre los estudiantes como lo arrojaron los resultados obtenidos de la investigación donde más del 60% de estudiantes tienen un nivel de uso alto de los mismos, se encuentra una relación significativa entre Ingresos mensuales y nivel de uso de una Tablet valores que se exponen en la tabla Nro. 5 de donde obtenemos un coeficiente de determinación de 2.959%.

Tomando en cuenta que en promedio en las instituciones universitarias privadas el costo por crédito de cada materia es alto y al ser la universidad en la que realizamos la investigación privada los recursos económicos son necesarios para matricularse. Los estudiantes de esta investigación pertenecen a entornos familiares con ingresos altos es lógico pensar en una relación directa entre Ingresos mensuales y en el semestre anterior en cuántas asignaturas se matriculó, pero nuestro análisis encontró una relación inversa entre estas variables con un valor de R de Pearson igual a -0.192, el nivel de significancia es de 0.000, como se observa en la tabla 6, el valor del coeficiente de determinación es 3.686 %, también se encuentra una relación indirecta entre Ingresos mensuales y en el semestre anterior en cuántas asignaturas aprobó con un valor de R de Pearson igual a -0.191, el nivel de significancia es de 0.000, como se observa en la tabla Nro. 5, el valor del coeficiente de determinación es 3.648 %, estas dos relaciones se pueden dar debido a la cantidad de alumnos que son becados o como la institución investigada lo dice quienes no cuenten con ingresos económicos familiares o pasen por alguna vicisitud financiera pueden adquirir descuentos y hasta becas para sus estudios.

5.1 Uso de las tecnologías en actividades académicas

5.11.1.4 Reducción de variables

Para reducir las variables obtenidas en este caso del literal Nro. 10, Nro. 11 y Nro. 14 del ANEXO 1, a un grupo de variables más manejables, se someten en primera instancia a una de las técnicas de minería de minería de datos como lo es análisis factorial, pero se comprueba que no cumplen con las condiciones explicadas por (De la Fuente, 2011) para este análisis que son:

- El valor del determinante de la matriz de correlación debe ser cercano a cero pero no necesariamente su valor es cero.
- El valor KMO (Kaiser, Meyer y Olkin) conocida como media de adecuación de la muestra debe ser cercana a uno.
- El test de esfericidad de Barlett debe ser significativa.
- La varianza debe explicar un valor alto de exactitud sin perder información

Para la clasificación de los estudiantes por el uso de las tecnologías en actividades académicas se obtuvo un valor de determinante de 0.231, una medida de adecuación de la muestra de 0.680, el test de esfericidad de Barlett no fue significativo y la varianza solo explica el 61.7 % de la información esto se lo puede observar en el ANEXO 5, datos similares salieron al aplicar análisis factorial al usos de las tecnologías en actividades de entretenimiento y uso de dispositivos electrónicos por lo que se determina que el método más factible para la investigación es el análisis clúster.

5.11.1.5 Clasificación de los estudiantes por el uso de tecnologías en actividades académicas.

Al concluir que el análisis factorial no era factible se aplica otra de las técnicas de minería de datos no supervisado que es el análisis clúster según (Weiss & Indurkha, 1998), procedemos a utilizar el programa SPSS 2009 para dividir la muestra en grupos. Se realizó clúster con 2,3 y 4 grupos (Ver ANEXO 6,7 y 8) para determinar con cuál de ellos podremos trabajar y representar de mejor manera. Del literal Nro. 10 de las preguntas realizadas a los estudiantes en el ANEXO 1, se toma las variables más representativas obtenidas por el análisis de estadística de dependencia central como lo es la mediana, que nos permitió identificar las preguntas con valores más representativos, estas preguntas se las ve en el literal 4.1.4 de este proyecto y son:

- ¿Aproximadamente cuántas consultas realiza a sus compañeros por mes?
- ¿Aproximadamente cuántos videos académicos mira en YouTube cada mes?
- ¿Aproximadamente cuántas horas chatea sobre temas académicos cada mes?
- ¿Aproximadamente cuántas horas busca información académica en internet cada mes?

Una vez realizado el análisis clúster se efectúa el análisis discriminante en el cual se utiliza como variable dependiente el número de grupos generado en el análisis clúster y como variables independientes las variables elegidas como más representativas, con este análisis podemos darnos cuenta el porcentaje de precisión para cada clasificación de clúster.

Tabla 6. Porcentaje de precisión de los clúster.

Nro. Clúster	Porcentaje de exactitud
2	99.7 %
3	98 %
4	98%

Fuente: Elaboración propia

La tabla Nro. 6, muestra los porcentajes de precisión una vez realizado el análisis discriminante a cada uno de los grupos obtenidos, ver ANEXO 9, el porcentaje de exactitud más alto tiene la clasificación para 2 grupos con 99.7%, pero para realizar una mejor explicación e interpretación se ha escogido la clasificación de 3 grupos que de igual manera tiene un porcentaje de exactitud alto de 98% la clasificación de tres grupos se la puede observar en la figura 15.

5.2.3 Resultado.

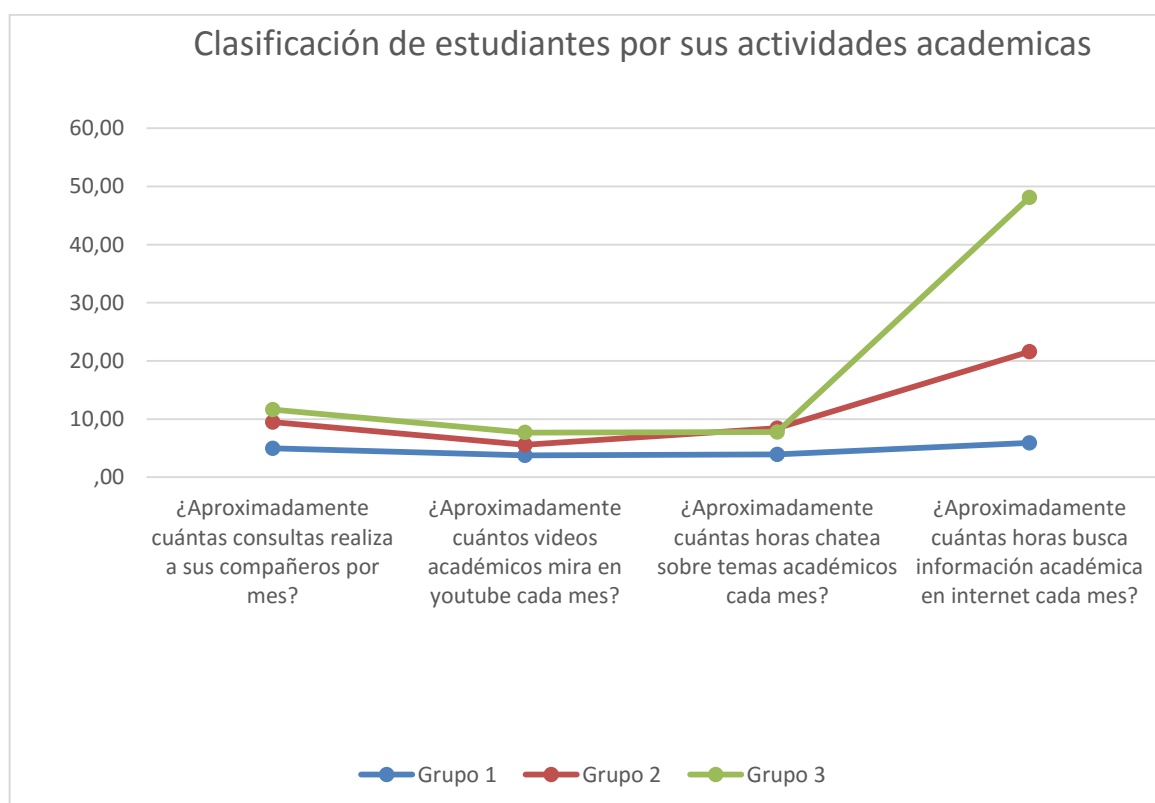


Figura 15. Análisis clúster en las actividades académicas

Fuente: Elaboración propia

El 66.4% de los estudiantes investigados pertenecen al Grupo Nro. 1 se lo denomina **BAJO**, tiene un nivel bajo de consultas a sus compañeros, mira pocos videos académicos, chatea

poco sobre temas académicos, y busca poca información académica en internet, es decir este grupo de estudiantes no usa las tecnologías e internet para realizar sus actividades académicas, no le toman o no entienden la verdadera importancia del uso de estos medios en sus actividades académicas.

El 27.2 % de los estudiantes investigados pertenecen al Grupo Nro. 2 se lo denomina **REGULAR**, este grupo realiza mayor cantidad de consultas a sus compañeros que el grupo bajo, en cuanto a los videos académicos que mira por mes no difiere mucho del grupo bajo, más bien tiene una similitud con el mismo, lo que si se empieza a diferenciar del grupo bajo es que chatea más sobre temas académicos, y busca mayor cantidad de horas información académica en internet, por lo que nos da a entender que este grupo se preocupa más por sus actividades académicas y su utilización de internet es mayor ya que aprecian las facilidades que les brinda esta herramienta.

El 6.4% de los estudiados investigados pertenecen al Grupo Nro. 3 se lo denomina **BUENO** el cual tiene un nivel más o menos similar al Grupo Nro. 2 de realizar consultas a sus compañeros y de ver videos académicos y de chatear sobre temas académicos, lo que si lo distingue verdadera mente de los demás grupos es que este busca mayor cantidad de información académica en internet, lo que hace suponer que de la muestra solo un poco porcentaje específicamente el 6.4% explota el internet en sus actividades académicas en la búsqueda de información para el posterior desarrollo de sus deberes.

5.3 Uso de las tecnologías en actividades de entretenimiento

5.11.1.6 Clasificación de los estudiantes por el uso de las tecnologías en actividades de entretenimiento

Para clasificar a los estudiantes por el uso de las tecnologías en actividades académicas se realiza de igual manera el análisis clúster con 2,3 y 4 grupos (Ver ANEXO 10,11 Y 12) para ver con cuál de ellos podremos trabajar de mejor manera. Del literal Nro. 11 del banco de preguntas, ver ANEXO 1, se ha tomado cuatro preguntas las que tenían valores más representativas para nuestro proyecto, las preguntas son:

- ¿Aproximadamente cuántas horas a la semana chatea por diversión?
- ¿Aproximadamente cuántas horas a la semana utiliza redes sociales?
- ¿Aproximadamente cuántas horas a la semana descarga música, videos y programas?
- ¿Aproximadamente cuántos videos para entretenimiento mira en YouTube cada semana?

Una vez que se realiza el análisis clúster se efectúa el análisis discriminante con el que se puede dar cuenta el porcentaje de precisión para cada clasificación de clúster.

Tabla 7. Porcentaje de precisión de los clúster.

Nro. Clúster	Porcentaje de exactitud
2	98 %
3	94.7 %
4	97.2 %

Fuente: Elaboración propia

La tabla 7, muestra los porcentajes de precisión una vez realizado el análisis discriminante a cada uno de los grupos obtenidos, ver ANEXO 13, el porcentaje de exactitud más alto tiene la clasificación para 2 grupos con 98 %, pero para poder explicar e interpretar de una manera más adecuada se ha escogido la clasificación de 3 grupos como se observa en la figura 16.

5.11.1.7 Resultado.

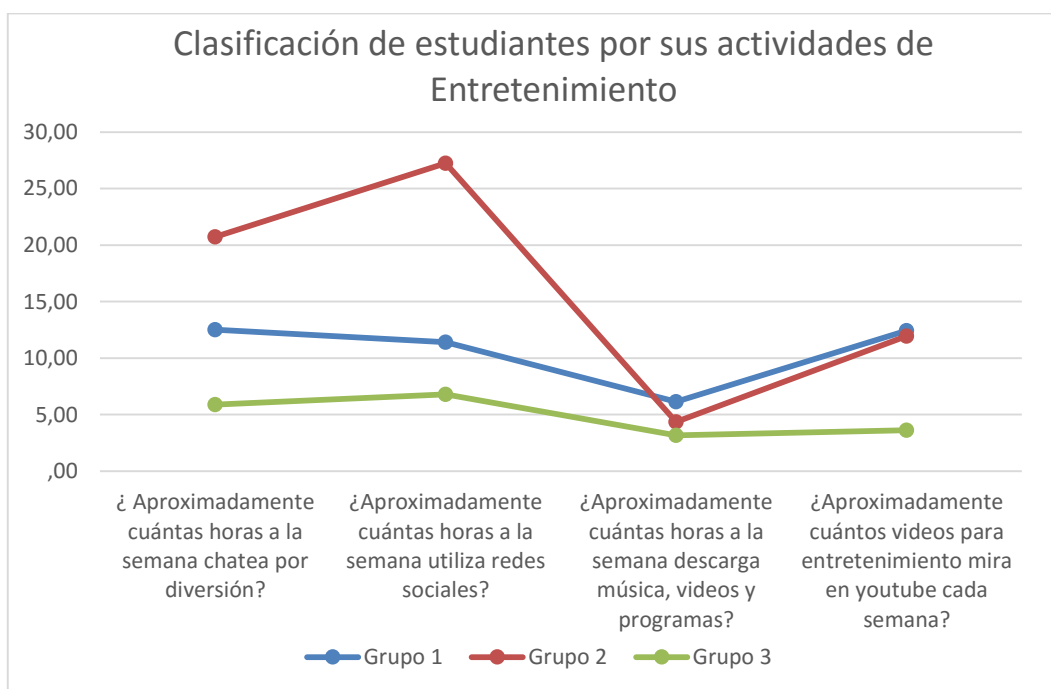


Figura 16. Análisis clúster en actividades de entretenimiento

Fuente: Elaboración propia

El 32.1 % de la muestra pertenece al Grupo Nro. 1 se lo denomina **MEDIO**, tiene un nivel medio de horas de chat por diversión, de uso de redes sociales y de descarga de contenidos en internet, este grupo tiene un nivel medio alto en lo que se refiere a la cantidad de videos que ve por semana para entretenimiento, en si este grupo usa el internet para su entretenimiento de una forma moderada, pero mayor que para actividades académicas.

El 24.7 % de la muestra pertenece al Grupo Nro. 2 se lo denomina **ACTIVO**, tiene un nivel alto de horas de chat y de uso de redes sociales, este grupo tiene valores similares en lo referente a la cantidad de videos de entretenimiento mira en YouTube por semana con el grupo medio y tiene un nivel bajo en lo que se refiere a la descarga de contenidos de internet incluso menor que el grupo medio, lo que se resalta de este grupo es que los alumnos utilizan el internet en mayor cantidad para las actividades de chat e interacción en las redes sociales, lo que nos da a entender que este grupo de alumnos se dedica más a estas actividades que a actividades académicas.

El 43.2 % de los encuestados perteneces al Grupo Nro. 3 se lo denomina **PASIVO** tiene un nivel bajo de horas de chat, uso de redes sociales, descarga de contenidos de internet, y la cantidad de videos que ve por semana para entretenimiento, a este grupo pertenece la mayor cantidad de alumnos lo que nos sugiere que los estudiantes pertenecientes a esta universidad utilizan poco el internet para actividades de entretenimiento, pero no se contrasta con el grupo bueno de uso de internet en actividades académicas al cual pertenece son el 6.4%.

5.3 Uso dispositivos electrónicos.

5.11.1.8 Clasificación de estudiantes por el uso de dispositivos electrónicos

Se aplica el mismo procedimiento aplicado en las sección 5.2.1 y 5.3.1 se realiza análisis clúster con 2,3 y 4 grupos, (Ver ANEXO 14, 15 y 16), para ver con cuál de ellos podremos trabajar de mejor manera. Del literal Nro. 14 del banco de preguntas aplicadas a los estudiantes, ver ANEXO 1, se tomó cuatro preguntas las que tienen los valores más representativos para realizar este proyecto, las preguntas son:

- Uso de Smartphone con cámara fotográfica y acceso a internet
- Uso de Teléfono móvil con acceso a internet
- Uso de Computador portátil
- Uso de Tablet (iPad; Galaxy Tab, Kindle, etc)

Cuando se termina el análisis clúster se realiza el análisis discriminante con el que se puede dar cuenta del porcentaje de precisión para cada clasificación de clúster.

Tabla 8. Porcentaje de precisión de los clúster.

Nro. Clúster	Porcentaje de exactitud
2	95.9 %
3	96.7 %
4	98.5 %

Fuente: Elaboración propia

La tabla 8, muestra los porcentajes de precisión una vez que se realiza el análisis discriminante a cada uno de los grupos obtenidos, ver ANEXO 17, el porcentaje de exactitud más alto tiene la clasificación para 4 grupos con 98.5 %, pero de igual manera se ha escogido la clasificación de 3 grupos para realizar la explicación y análisis de grupos como se observa en la figura 17.

5.11.1.9 Resultado.

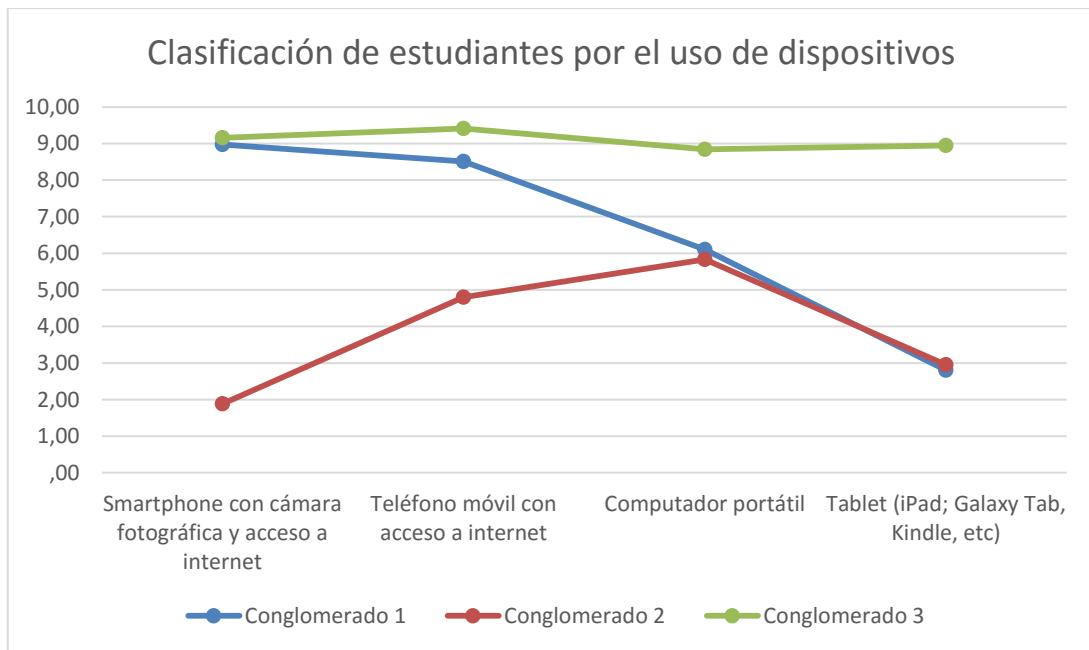


Figura 17. Análisis clúster uso de dispositivos electrónicos
Fuente: Elaboración propia

El 27.7 % de los encuestados perteneces al Conglomerado Nro. 1 al cual se lo denomina **NORMAL** tiene un nivel alto en el uso de un Smartphone y de teléfono con acceso a internet, tienen un nivel medio de uso de un computador portátil y un nivel bajo de uso de una Tablet, lo que no se aparta de la realidad de nuestra investigación pues el uso de teléfonos celulares en la actualidad es parte importante de la vida diaria estudiantil tanto para actividades académicas como de entretenimiento, con la aparición de los Smartphone se facilita la búsqueda de información e interacción social.

El 15 % de los encuestados perteneces al Conglomerado Nro. 2 al cual se lo denomina **IRREGULAR** porque al contrario del grupo Normal tiene un nivel bajo en el uso de un Smartphone y de una Tablet, tiene un nivel medio del uso de un teléfono con acceso a internet, y uso de un computador portátil.

El 57.3 % de los encuestados perteneces al Conglomerado Nro. 3 al cual se lo denomina **CONSTANTE** tiene un tiene un nivel alto en el uso de un Smartphone, de un teléfono con acceso a internet, de una Tablet, y uso de un computador portátil, en este grupo se encuentra la mayor cantidad de la muestra, al pertenecer la mayor cantidad de la población estudiada a un grupo con ingresos económicos altos es lógico que cuenten con aparatos electrónicos de última tecnología para el desarrollo de sus actividades tanto de entretenimiento como académicas.

5.4 Rendimiento académico.

En el trabajo realizado por (Hernández, 2002) se define el fracaso escolar como “el problema de las mil causas” debido a la innumerable cantidad de posible de causas o factores de tipo tanto personal, académicas, físicas, económicas, familiares, sociales, institucionales, pedagógicas, etc., que pueden tener influenciar en el fracaso o abandono de los estudiantes como lo menciona (Vera et al., 2012).

5.11.1.10 Clasificación de los estudiantes por su rendimiento académico

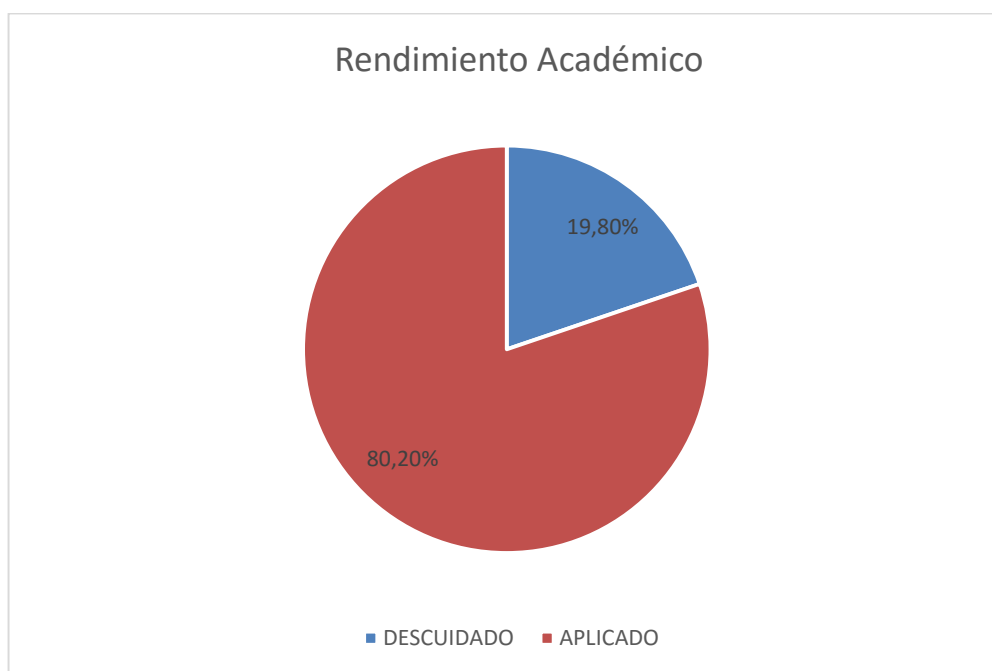


Figura 18. Rendimiento académico

Fuente: Elaboración propia

Se clasifica a los estudiantes en dos grupos como se lo observa en la figura 18, en el primer grupo están quienes no han reprobado ni una materia y se los denomina como el grupo de **APLICADOS** y a quienes han perdido materias se los ha denominado como alumnos **DESCUIDADOS**, los alumnos **APLICADOS** corresponden al 80.2 % del total de la muestra mientras que los alumnos **DESCUIDADOS** corresponden al 19.8 % del total de la muestra y

están conformados por aquellos que han perdido 1, 2 y 3 materias como se observa en la tabla Nro. 9.

Tabla 9. Rendimiento Académico

RENDIMIENTO ACADEMICO				
	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
0	315	80.2	80.2	80.2
1	52	13.2	13.2	93.4
2	21	5.3	5.3	98.7
3	5	1.3	1.3	100.0
Total	393	100.0	100.0	

Fuente: Elaboración propia

5.5 Comprobación de hipótesis

5.11.1.11 Comprobación de hipótesis 1: El nivel de ingresos determina como se utiliza internet para el aprendizaje

Para comprobar esta hipótesis primero vamos a determinar la hipótesis nula y alternativa que son:

Ho: El nivel de ingresos establece como se utiliza internet para el aprendizaje

H1: El nivel de ingresos no establece como se utiliza internet para el aprendizaje

Para esto se va a tener en cuenta que el nivel socio económico de la muestra es alto el 73% de los estudiantes encuestados tiene ingresos superiores a 1000 dólares, Figura 4, otro factor a tomar en cuenta es que se clasifica a los estudiantes y el uso de la tecnología en actividades académicas en tres grupos que son Bajo, Regular y Bueno.

Para el análisis de esta hipótesis se va a tomar en cuenta a los ingresos mensuales como variable independiente mientras que los clúster académicos van a ser tomado como variable dependiente, para este análisis se va a utilizar una de las técnicas de minería de datos supervisados o predictivos que es la regresión logística multinomial.

La varianza que explica el modelo mencionada a través del valor R² Nagelkerke llega a 0,027 (2.7 %). Los resultados de la prueba de Wald (ver Tabla 10) indican que en los dos modelos obtenidos todos los niveles de la variable ingreso no son significativos ($p > 0,05$) con lo que rechazamos al Ho.

En la tabla de Wald también encontramos los Odds ratio y sus respectivos intervalos de confianza. Interpretando estos resultados tenemos que el odds ratio de pertenecer al perfil académico BUENO con respecto al perfil BAJO:

- Disminuye 3.91 veces cuando el estudiante pertenece al nivel de ingreso 1 respecto al nivel 5 (OR = 0.256, (IC 95% 0.032 – 2.084), $p=0,203$).
- El modelo es 1,144 veces mayor cuando pertenece al nivel de ingreso 2 respecto al nivel 5 (OR = 1,144, (IC 95% 0,390 – 3,360), $p=0,806$).

- Disminuye 1.555 veces cuando el estudiante pertenece al nivel de ingreso 3 respecto al nivel 5 (OR = 0,643, (IC 95% 0,192 – 2,159), p=0,475).
- Disminuye 1.615 veces cuando pertenece al nivel de ingreso 4 respecto al nivel 5 (OR = 0,619, (IC 95% 0,185 – 2,075), p=0,437).

En el segundo modelo, el OR de pertenecer al perfil académico REGULAR con respecto al perfil BAJO da los siguiente resultados

- Disminuye 2.02 veces cuando el estudiante pertenece al nivel de ingreso 1 respecto al nivel 5 (OR = 0.496, (IC 95% 0.189 – 1.302), p=0,154)
- Aumenta 1,107 veces cuando pertenece al nivel de ingreso 2 respecto al nivel 5 (OR = 1,107, (IC 95% 0,556 – 2,203), p=0,772).
- Aumenta 1,452 veces cuando el estudiante pertenece al nivel de ingreso 3 respecto al nivel 5 (OR = 1,452, (IC 95% 0,782 – 2,697), p=0,238).
- Se incrementa 1,198 veces cuando pertenece al nivel de ingreso 4 respecto al nivel 5 (OR = 1,198, (IC 95% 0,635 – 2,260), p=0,577).

Tabla 10 Coeficiente modelo de regresión

		Estimaciones de los parámetros					Intervalo de confianza al 95% para Exp(B)		
CLUS_ACA ^a		B	Error típ.	Wald	gl	Sig.	Exp(B)	Límite inferior	Límite superior
BUENO	Intersección	-2,104	,335	39,461	1	,000			
	[ing=1.00]	-1,362	1,069	1,621	1	,203	,256	,032	2,084
	[ing=2.00]	,135	,550	,060	1	,806	1,144	,390	3,360
	[ing=3.00]	-,441	,618	,510	1	,475	,643	,192	2,159
	[ing=4.00]	-,480	,617	,604	1	,437	,619	,185	2,075
	[ing=5.00]	0 ^b	.	.	0
REGULAR	Intersección	-,973	,211	21,285	1	,000			
	[ing=1.00]	-,701	,492	2,029	1	,154	,496	,189	1,302
	[ing=2.00]	,102	,351	,084	1	,772	1,107	,556	2,203
	[ing=3.00]	,373	,316	1,395	1	,238	1,452	,782	2,697
	[ing=4.00]	,180	,324	,310	1	,577	1,198	,635	2,260
	[ing=5.00]	0 ^b	.	.	0

a. La categoría de referencia es: BAJO.

b. Este parámetro se ha establecido a cero porque es redundante.

Fuente: Elaboración propia

Con los resultados obtenidos podemos rechazar la hipótesis nula y se va a aceptar la hipótesis alternativa que dice que el nivel de ingreso no determina con se utiliza el internet en el aprendizaje.

5.11.1.12 Comprobación hipótesis 2: El nivel de ingresos determina como se utiliza internet para entretenimiento.

Al igual que en la primera hipótesis se va a determinar primeramente la hipótesis nula y alternativa que son:

Ho: El nivel de ingresos establece como se utiliza internet para el aprendizaje

H1: El nivel de ingresos no establece como se utiliza internet para el aprendizaje

En esta hipótesis también es necesario tener en cuenta que el nivel socio económico de la muestra es alto pues el 73% de los estudiantes encuestados tiene ingresos económicos familiares superiores a 1000 dólares, Figura 4, otro factor que se va a tomar en cuenta para comprobar esta hipótesis es la clasificación de los estudiantes y su uso de la tecnología en actividades de entretenimiento que se la realiza en tres grupos que son Medio, Activo y Pasivo. Para el análisis de esta hipótesis se va a tomar en cuenta que los ingresos mensuales como variable Independiente mientras que a los clúster de entretenimiento se los va a tomar como variables dependientes, al igual que para el análisis anterior se va a utilizar una regresión logística multinomial.

La varianza que explica el modelo expuesta a través del valor R² Nagelkerke llega a 0,027 (2.7 %). Los resultados de la prueba de Wald (Tabla Nro. 11) presentan solo 2 valores de nivel de significancia aceptable para variables de ingreso ($p < 0,05$).

En el primer modelo, el OR de pertenecer al perfil de entretenimiento ACTIVO con respecto al perfil PASIVO nos predice lo siguiente:

- Disminuye 6.33 veces cuando el estudiante pertenece al nivel de ingreso 1 respecto al nivel 5 (OR = 0,158, (IC 95% 0.044 – 0.570), **p=0,005**).
- Disminuye 1.31 veces mayor cuando pertenece al nivel de ingreso 2 respecto al nivel 5 (OR = 0,759, (IC 95% 0,363 – 1,586), p=0,463).
- Disminuye 1,35 veces cuando el estudiante pertenece al nivel de ingreso 3 respecto al nivel 5 (OR = 0,740, (IC 95% 0,377 – 1,454), p=0,383).
- Disminuye 1.41 veces mayor cuando pertenece al nivel de ingreso 4 respecto al nivel 5 (OR = 0,711, (IC 95% 0,342 – 1,479), p=0,362).

En el segundo modelo, el OR de pertenecer al perfil académico MEDIO con respecto al perfil PASIVO nos dice lo siguiente:

- Disminuye 2.141 veces cuando el estudiante pertenece al nivel de ingreso 1 respecto al nivel 5 (OR = 0.467, (IC 95% 0.208 – 1.049), p=0,065).
- Disminuye 1.689 veces cuando pertenece al nivel de ingreso 2 respecto al nivel 5 (OR = 0,592, (IC 95% 0,291 – 1,203), p=0,147).
- Disminuye 2.44 veces cuando el estudiante pertenece al nivel de ingreso 3 respecto al nivel 5 (OR = 0,410, (IC 95% 0,205 – 0,820), **p=0,012**).
- Disminuye 1.10 veces cuando pertenece al nivel de ingreso 4 respecto al nivel 5 (OR = 0,906, (IC 95% 0,475 – 1,727), p=0,763).

Tabla 11. Coeficiente del modelo de regresión

Estimaciones de los parámetros									
CLUS_ENTRE ^a	B	Error típ.	Wald	gl	Sig.	Exp(B)	Intervalo de confianza al 95% para Exp(B)		
							Límite inferior	Límite superior	
ACTIVO	Intersección	-,235	,229	1,047	1	,306			
	[ing=1.00]	-1,845	,654	7,956	1	,005	,158	,044	,570
	[ing=2.00]	-,276	,376	,538	1	,463	,759	,363	1,586
	[ing=3.00]	-,301	,345	,761	1	,383	,740	,377	1,454
	[ing=4.00]	-,341	,373	,831	1	,362	,711	,342	1,479
	[ing=5.00]	0 ^b	.	.	0
MEDIO	Intersección	,067	,212	,101	1	,751			
	[ing=1.00]	-,761	,412	3,403	1	,065	,467	,208	1,049
	[ing=2.00]	-,524	,362	2,098	1	,147	,592	,291	1,203
	[ing=3.00]	-,891	,353	6,349	1	,012	,410	,205	,820
	[ing=4.00]	-,099	,329	,091	1	,763	,906	,475	1,727
	[ing=5.00]	0 ^b	.	.	0

a. La categoría de referencia es: PASIVO.

b. Este parámetro se ha establecido a cero porque es redundante.

Fuente: Elaboración propia

Es decir que solo se va a cumplir la Ho cuando el estudiante pertenece al nivel de ingreso 1, es decir en ingresos familiares son hasta 350 dólares, y al nivel de ingreso 3, es decir en ingresos familiares son hasta 1000 dólares, en estos dos casos el nivel de ingresos determinara como se utiliza el internet para actividades de entretenimiento, para los demás casos en esta investigación se rechaza la Ho y se acepta H1.

5.11.1.13 Comprobación hipótesis 3: El uso de la tecnología en el aprendizaje incide en el rendimiento académico.

Se determina primeramente la H_0 y la H_1 que son:

H_0 : El uso de la tecnología en el aprendizaje incide en el rendimiento académico.

H_1 : El uso de la tecnología en el aprendizaje no incide en el rendimiento académico.

Para comprobar si el uso de la tecnología en el aprendizaje incide en el rendimiento académico se va a tomar en cuenta como variable independiente a los clúster académicos mientras que como variables dependientes al rendimiento académico, para esto se elabora un modelo de regresión logística binaria en el que el valor de la prueba chi cuadrado del logaritmo de la verosimilitud no tiene un valor significativo ($X^2=12,909$, $p=0,865$), lo que indica que se trata de un modelo no significativo ($p > 0,05$); el valor de R^2 de Negelkerke llega al 1,0%.

Los resultados del test de Wald indican que cada una de las categorías del perfil académico tiene coeficientes distintos de cero, la tabla Nro. 12 también presenta los OR y su intervalo de confianza

El OR de los estudiantes con rendimiento aplicado respecto a los estudiantes con perfil académico descuidado nos da los siguientes resultados:

- Es 1.077 veces mayor cuando el alumno pertenece al perfil académico bajo en comparación con el perfil académico regular (OR=1.077, (IC 95% 0,550 – 2,111), $p=0,828$)
- Disminuye 1.266 (1/.790) veces cuando el alumno pertenece al perfil académico bajo en comparación con el perfil académico regular (OR=1.266, (IC 95% 0,236 – 2,645), $p=0,703$).

Con los resultados obtenidos se rechaza la H_0 y se acepta H_1 que nos dice que el uso de la tecnología en el aprendizaje no incide en el rendimiento académico.

Tabla 12. Coeficiente del modelo de regresión

		Estimaciones de los parámetros					Intervalo de confianza al 95% para Exp(B)		
Rendimiento_cat ^a		B	Error típ.	Wald	gl	Sig.	Exp(B)	Límite inferior	Límite superior
aplicado	Intersección	1,894	,287	43,629	1	,000			
	[CLUS_ACA=BAJO]	,074	,343	,047	1	,828	1,077	,550	2,111
	[CLUS_ACA=BUENO]	-,235	,616	,146	1	,703	,790	,236	2,645
	[CLUS_ACA=REGULAR]	0 ^b	.	.	0

a. La categoría de referencia es: descuidado.

b. Este parámetro se ha establecido a cero porque es redundante.

Fuente. Elaboración propia

5.11.1.14 Comprobación hipótesis 4: El uso de la tecnología para entretenimiento incide en el rendimiento académico.

Para comprobar esta hipótesis de igual manera se va a determinar primeramente la Ho y la H1 que son:

Ho: El uso de la tecnología para entretenimiento incide en el rendimiento académico.

H1: El uso de la tecnología para entretenimiento no incide en el rendimiento académico.

En el presente análisis se va a tomar en cuenta como variable independiente a los clúster de entretenimiento mientras que como variables dependientes al rendimiento académico y para esto se elaboró un modelo de regresión logística binaria en el que el valor de la prueba chi cuadrado del logaritmo de la verosimilitud no tiene un valor significativo ($X^2=13,552$, $p=0,390$), lo que indica que se trata de un modelo no significativo ($p > 0,05$); el valor de R2 de Nagelkerke llega al 0.9 %.

Los resultados del test de Wald indican que cada una de las categorías del perfil académico tiene coeficientes distintos de cero, la tabla nro. 13 también presenta los OR y su intervalo de confianza

El OR de los estudiantes con rendimiento aplicado respecto al perfil de los estudiantes con perfil académico descuidado nos dice lo siguiente:

- Es 1.667(1/0.600) veces menor cuando el alumno pertenece al perfil de entretenimiento activo en comparación con el perfil académico pasivo (OR=0,600, (IC 95% 0,290 – 1,238), $p=0,167$).
- Es 1.228 (1/.814) veces menor cuando el alumno pertenece al perfil de entretenimiento

medio respecto al perfil de entretenimiento pasivo (OR=0.814, (IC 95% 0,398 – 1,667), p=0,574).

Tabla 13. Coeficiente del modelo de regresión

		Estimaciones de los parámetros					Intervalo de confianza al 95% para Exp(B)		
Rendimiento_cat ^a		B	Error típ.	Wald	gl	Sig.	Exp(B)	Límite inferior	Límite superior
aplicado	Intersección	2,134	,249	73,258	1	,000			
	[CLUS_ENTRE=ACTIVO]	-,512	,370	1,911	1	,167	,600	,290	1,238
	[CLUS_ENTRE=MEDIO]	-,206	,366	,316	1	,574	,814	,398	1,667
	[CLUS_ENTRE=PASIVO]	0 ^b	.	.	0

a. La categoría de referencia es: descuidado.

b. Este parámetro se ha establecido a cero porque es redundante.

Fuente elaboración propia

Con los resultados obtenidos se rechaza la Ho y se acepta H1 que nos dice que el uso de la tecnología para entretenimiento no incide en el rendimiento académico.

CAPITULO VI
OBSERVACIONES FINALES Y CONCLUSIONES

6.1 Observaciones finales

En este capítulo se analizan los resultados obtenidos en la comprobación de hipótesis, se dilucida las relaciones e incidencias que tienen las variables como la edad, el sexo, los ingresos económicos, sobre el uso del internet y tecnologías en actividades tanto académicas como de entretenimiento y como estas afectan al rendimiento académico estudiantil.

Para el análisis de los resultados hemos tomado en cuenta los siguientes puntos:

- **Entorno estudiantil**

El nivel económico según los datos obtenidos en la Universidad Católica Santiago de Guayaquil son altos, el 73% del total de encuestados poseen un ingreso superior a 1000 dólares, por lo que se analiza la relación de estos ingresos con el uso de internet tanto en actividades académico como de entretenimiento.

A pesar que los estudiantes investigados poseen un ingreso familiar alto y tienen la facilidad de adquirir planes de datos y conectarse desde una red móvil; esta investigación demuestra que solo el 23.9% de los estudiantes se conecta desde la misma y que la mayoría de ellos se conectan desde la casa con un 56.7% del total.

Los ingresos económicos familiares no determinaron, en nuestra investigación, como los estudiantes utilizan el internet para actividades académicas y de entretenimiento. El uso de internet en actividades académicas varía, se demostró que el 6.4 % pertenece al grupo BUENO que son aquellos estudiantes que utilizan más el internet para actividades académicas; de igual forma aunque el 85 % de la muestra se conecte 7 días a la semana con un promedio de 7.7 horas de conexión diaria, el 43.2% pertenecen al grupo PASIVO, es decir, aquellos estudiantes con poco uso de internet para actividades de entretenimiento. Con esto se comprobó que no se cumple la hipótesis que dice que el nivel de ingresos determina como se utiliza internet para el aprendizaje y además se rechazó la hipótesis que sostiene que el nivel de ingresos determina como se utiliza internet para actividades de entretenimiento.

Al encontrar que más del 50% de los estudiantes se conectan desde su hogar, los resultados obtenidos en cuanto al manejo de internet concuerdan con lo expuesto por (Hargittai, 2010), es decir, quienes cuentan con una conexión desde el hogar tienden a tener mayor conocimiento y destrezas en el manejo de internet.

El alto nivel de conocimiento en el manejo de internet también se debe a que el 85% de la población universitaria se conecta 7 días a la semana con una media de conexión diaria de 7.7 horas, lo que permite a cada uno de ellos estar peritos en el uso de las tecnologías y de internet para cualquier actividad sea académica o de entretenimiento.

- **Uso del Internet en actividades académicas**

Los estudiantes investigados tienen una media de 3 ingresos semanales a la plataforma virtual de la universidad, así como una media de 7 consultas a sus compañeros, una media de 5 tanto en videos académicos vistos al mes, como en horas de chat sobre temas académicos y una media de 13 horas en la búsqueda de información académica al mes.

Una vez realizada la clasificación de estudiantes por sus hábitos de uso de internet en actividades académicas se obtuvieron tres grupos de estudiantes denominados BAJO, REGULAR Y BUENO.

Primeramente se evidencia el mayor porcentaje de la muestra es decir el 66.4% pertenece al grupo denominado BAJO, al que pertenecen los estudiantes que realizan un bajo número de consultas a sus compañeros, observan pocos videos académicos, tienen un nivel bajo de horas de chat sobre temas académicos y busca poca información en internet para el desarrollo de sus actividades académicas.

El grupo de estudiantes tanto REGULAR como BUENO, tienen características similares en lo concerniente a consultas a sus compañeros, al número de videos académicos vistos al mes y al número de horas que chatea mensualmente sobre temas académicos, lo que diferencia al grupo REGULAR y BUENO son las horas de búsqueda de información para actividades académicas siendo el grupo BUENO quien tiene mayor cantidad de horas de búsqueda de información.

Se encuentra que el 94.7 % de los encuestados sostiene que internet le permite realizar de forma rápida y con menor esfuerzo los trabajos y un 89.3% de los encuestados tiene un porcentaje alto de aceptación de la información de internet; lo que no se contrasta con lo encontrado en esta investigación, debido a que solo el 6.4% pertenece al grupo de estudiantes con alto uso de internet en actividades académicas.

Es interesante saber que el 80.2 % de la muestra pertenece al grupo de estudiantes APLICADOS y el 19.8 % pertenece a los estudiantes DESCUIDADOS. Con estos datos se evidencia que el uso de internet en actividades académicas no influye en el rendimiento académico; los estudiantes pertenecientes al grupo BAJO lo conforman más del 50 % de la muestra, por lo que se rechazó la hipótesis que sostiene que el uso de las tecnologías en el aprendizaje inciden en el rendimiento académico.

- **Uso del Internet en Actividades de entretenimiento**

Una vez realizada la clasificación de los estudiantes por el nivel de uso de internet en actividades de entretenimiento se ha obtenido tres grupos denominados PASIVO, MEDIO y ACTIVO.

El grupo denominado PASIVO representa el mayor número de estudiantes con un 43.2%, el uso de internet para actividades de entretenimiento es bajo, el número de horas de chat por diversión se contrasta con las pocas horas que utilizan redes sociales, el número de horas que descarga contenidos de internet es similar al número de videos que ve por entretenimiento en la red.

El grupo denominado MEDIO representa el 32.1 % de la muestra y tiene mayor número de horas de chat que de ingreso a redes sociales, lo que demuestra que el chat lo realiza por otros medios como mensajes de texto o WhatsApp. El número de descargas de contenidos es menor que el número de videos que observa por entretenimiento.

El grupo denominado ACTIVO representa el 24.7 % de los encuestados, este grupo tienen un número muy elevado de ingreso a redes sociales, por lo que las horas de chat son elevadas, un dato singular es que la descarga de contenidos es ligeramente mayor que el grupo PASIVO y es menor que el grupo MEDIO, además el número de videos observados por entretenimiento es igual al grupo MEDIO, siendo la única diferencia con los demás grupos de estudiantes es que estos son más preocupados de las relaciones sociales.

Hay que tomar en cuenta que 56.8 % del total de la muestra sumando el grupo MEDIO y ACTIVO usan en mayor cantidad el internet para actividades de entretenimiento y estas no influye en el rendimiento académico, ya que el 80.2% pertenece al grupo APLICADO, es decir, quienes no han reprobado materias, de esta forma se comprueba que no se cumple la hipótesis que dice que el uso de las tecnologías en entretenimiento incide en el rendimiento académico.

El porcentaje de uso de dispositivos es elevado, tomando como referencia aquellos estudiantes que ponderaron sus respuestas en la escala de 5 hasta 10, se encontró que el porcentaje de uso de un Smartphone es del 85.5%, de un teléfono móvil con acceso a internet es del 89.8%, de un computador portátil es del 82.4 % y de una Tablet es del 67.4%.

Tomando en cuenta estos valores es lógico que el grupo medio y el grupo activo representen el 56.8% del total de los encuestados, ya que los niveles de usos de dispositivos son altos especialmente en los dispositivos electrónicos utilizados para el entretenimiento.

6.2 Conclusiones generales

En base a los resultados obtenidos del análisis podemos concluir lo siguiente:

- En este proyecto se ha demostrado que los ingresos mensuales no influyen en el uso del internet en actividades académicas, a pesar que el 73 % de la muestra tienen ingresos altos, solo el 6.4% de la misma utiliza el internet en actividades académicas.
- Se determina que los ingresos familiares influyen en el uso de internet en actividades de entretenimiento cuando los estudiantes poseen ingresos de 350 dólares y de 1000 dólares.
- Se determina que el uso de internet en actividades académicas no incide en el rendimiento académico, se encontró que existe un porcentaje bajo de 6.4% de estudiantes que utiliza con frecuencia el internet para actividades académicas, a pesar de estos resultados, un 80.2% de estudiantes no han reprobado materias.
- El uso de internet en actividades de entretenimiento no influye en el rendimiento académico, se comprobó que un 56.8% de estudiantes investigados posee un nivel medio alto de uso de internet para actividades de entretenimiento pero no influye en su rendimiento académico porque existe un porcentaje alto de alumnos que no han reprobado materias.

ANEXOS

ANEXO 1

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.

1. Responda la siguiente pregunta	
¿En qué universidad estudia?	

2. Responda la siguiente pregunta	
¿Que carrera estudia?	

3. Responda la siguiente pregunta	
¿Cuál es su edad?	

4. Responda la siguiente pregunta	Hombre	Mujer
¿Cuál es su género?	()	()

5. Los ingresos mensuales de su familia son de:	
Hasta 350 dólares	()
Hasta 600 dólares	()
Hasta 1.000 dólares	()
Hasta 1.500 dólares	()
Más de 1.500 dólares	()

6. ¿Desde dónde se conecta habitualmente a Internet? (escoja solo una opción)	
Desde la casa	()
Desde un cyber café	()
Desde el trabajo	()
Desde la Universidad	()
Desde una red móvil (movistar, claro, cnt)	()

7. Responda la siguiente pregunta	1	2	3	4	5	6	7
De 1 a 7, ¿cuántos días a la semana se conecta Internet?	()	()	()	()	()	()	()

8. Responda las siguientes preguntas	1	2	3	4	5	6	7	8	9	10
De 1 a 10 su nivel de conocimientos en el manejo de Internet es:	()	()	()	()	()	()	()	()	()	()

9. Responda las siguientes preguntas	
¿Aproximadamente cuántas horas se conecta cada día?	(____)
¿Hace cuántos años se conecta a Internet?	(____)

10. En lo referente a las asignaturas en las que está matriculado	
¿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?	(____)

11. En lo referente al entretenimiento y diversión en internet	
¿ 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?	(____)

12. Responda las siguientes preguntas	
¿ Aproximadamente cuántos seguidores tiene en twitter?	(____)
¿ Cuántos amigos tiene en facebook?	(____)
¿ Cuántos contactos tiene en LinkedIn?	(____)

13. Responda con una X en SI o NO a las siguientes preguntas	SI	No
Tiene un blog	()	()
Tiene cuenta en youtube	()	()
Tiene cuenta en www.del.icio.us	()	()

14. ¿Cuál es su nivel de uso de los siguientes dispositivos? (1 significa no usar y 10 significa utilizar al máximo)	1	2	3	4	5	6	7	8	9	10
Smartphone con cámara fotográfica y acceso a internet	()	()	()	()	()	()	()	()	()	()
Teléfono móvil con acceso a internet	()	()	()	()	()	()	()	()	()	()
Teléfono móvil sin acceso a internet	()	()	()	()	()	()	()	()	()	()
Computador portátil	()	()	()	()	()	()	()	()	()	()
Tablet (iPad; Galaxy Tab, Kindle, etc)	()	()	()	()	()	()	()	()	()	()
Cámara digital	()	()	()	()	()	()	()	()	()	()

iPod / MP3 Player	()	()	()	()	()	()	()	()	()	()
-------------------	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

15. De 1 a 10 valore los siguientes aspectos (1 significa no estar de acuerdo y 10 estar completamente de acuerdo)

	1	2	3	4	5	6	7	8	9	10
Internet le permite elaborar los trabajos más rápido y con menos esfuerzo	()	()	()	()	()	()	()	()	()	()
Usted confía en la información de internet para realizar sus tareas	()	()	()	()	()	()	()	()	()	()
Internet le permite prescindir de la Biblioteca	()	()	()	()	()	()	()	()	()	()
Internet facilita el proceso de aprendizaje	()	()	()	()	()	()	()	()	()	()
Internet le permite mejorar sus calificaciones	()	()	()	()	()	()	()	()	()	()
Usted presenta trabajos académicos copiados desde Internet	()	()	()	()	()	()	()	()	()	()

16. Responda las siguientes preguntas referentes a sus profesores. (Se recomienda evaluar de forma general a todos sus profesores)

	SI	NO	A veces
Su profesor ingresa a la plataforma virtual	()	()	()
Contesta sus consultas por correo electrónico	()	()	()
Chatea con usted eventualmente sobre aspectos académicos	()	()	()
Su profesor comenta en redes sociales sobre temas académicos	()	()	()
Le sube materiales digitales a la plataforma virtual	()	()	()
Le recomienda recursos digitales de la biblioteca virtual	()	()	()
Le recomienda videos sobre temas académicos	()	()	()
Le plantea cuestionarios o evaluaciones en la plataforma virtual	()	()	()
Le plantea foros virtuales	()	()	()
Su profesor tiene una página web, blog o perfil de facebook	()	()	()
Su profesor tiene cuenta de twitter	()	()	()

17. Responda las siguientes preguntas:

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

ANEXO 2

Tabla de contingencia ¿Cuál es su género? * ¿Desde dónde se conecta habitualmente a Internet? (escoja solo una opción)

			¿Desde dónde se conecta habitualmente a Internet? (escoja solo una opción)					Total
			Desde la casa	Desde un cyber café	Desde el trabajo	Desde la Universidad	Desde una red móvil (movistar, claro, cnt)	
¿Cuál es su género?	Hombre	Recuento	92	2	1	26	27	148
		% dentro de ¿Cuál es su género?	62,2%	1,4%	,7%	17,6%	18,2%	100,0%
	Mujer	Recuento	131	3	1	43	67	245
		% dentro de ¿Cuál es su género?	53,5%	1,2%	,4%	17,6%	27,3%	100,0%
Total		Recuento	223	5	2	69	94	393
		% dentro de ¿Cuál es su género?	56,7%	1,3%	,5%	17,6%	23,9%	100,0%

ANEXO 3

Pruebas de chi-cuadrado

	Valor	Gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	4,567 ^a	4	,335
Razón de verosimilitudes	4,675	4	,322
Asociación lineal por lineal	3,699	1	,054
N de casos válidos	393		

a. 4 casillas (40.0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es .75.

ANEXO 4

Medidas simétricas

		Valor	Error típ. asint. ^a	T aproximada ^b	Sig. aproximada
Nominal por nominal	Phi	.745			.134
	V de Cramer	.199			.134
	Coeficiente de contingencia	.597			.134
Ordinal por ordinal	Tau-b de Kendall	-.092	.035	-2.594	.009
	Tau-c de Kendall	-.084	.032	-2.594	.009
	Correlación de Spearman	-.123	.047	-2.441	.015 ^c
Intervalo por intervalo	R de Pearson	-.106	.043	-2.115	.035 ^c
N de casos válidos		393			

a. Asumiendo la hipótesis alternativa.

b. Empleando el error típico asintótico basado en la hipótesis nula.

c. Basada en la aproximación normal.

ANEXO 5

Matriz de correlaciones^a

	¿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?
Correlación	1,000	,330	,151	,301	,067	,100	-,044	,026	,162	,112
¿Cuántas veces a la semana ingresa a la plataforma virtual de su universidad?										
¿Aproximadamente cuántas consultas les hace a sus profesores cada mes?	,330	1,000	,454	,273	,198	,013	,053	,169	,174	,269
¿Aproximadamente cuántas consultas realiza a sus compañeros por mes?	,151	,454	1,000	,218	,210	,011	-,009	,185	,350	,245
¿Aproximadamente cuántos recursos educativos descarga de la plataforma virtual cada mes?	,301	,273	,218	1,000	,312	,075	,145	,209	,175	,152
¿Aproximadamente cuántos videos académicos mira en youtube cada mes?	,067	,198	,210	,312	1,000	,113	,276	,182	,279	,268
¿Aproximadamente en cuántos foros virtuales participa cada mes?	,100	,013	,011	,075	,113	1,000	,116	,077	-,042	,190
¿Aproximadamente cuántos post o tweets sobre temas académicos realiza en las redes sociales por mes?	-,044	,053	-,009	,145	,276	,116	1,000	,271	,052	,042
¿Aproximadamente cuántas horas chatea sobre temas académicos cada mes?	,026	,169	,185	,209	,182	,077	,271	1,000	,295	,336
¿Aproximadamente cuántas horas busca información académica en internet cada mes?	,162	,174	,350	,175	,279	-,042	,052	,295	1,000	,325
¿Aproximadamente cuántas horas utiliza la biblioteca virtual de la universidad cada mes?	,112	,269	,245	,152	,268	,190	,042	,336	,325	1,000

a. Determinante = ,231

KMO y prueba de Bartlett

Medida de adecuación muestral de Kaiser-Meyer-Olkin.		,680
Prueba de esfericidad de Chi-cuadrado aproximado		568,552
Bartlett	GI	45
	Sig.	,000

Varianza total explicada

Componente	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción			Suma de las saturaciones al cuadrado de la rotación		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	2,702	27,023	27,023	2,702	27,023	27,023	1,893	18,934	18,934
2	1,340	13,400	40,422	1,340	13,400	40,422	1,691	16,910	35,844
3	1,123	11,226	51,648	1,123	11,226	51,648	1,482	14,824	50,668
4	1,005	10,055	61,703	1,005	10,055	61,703	1,104	11,035	61,703
5	,811	8,114	69,817						
6	,810	8,095	77,913						
7	,670	6,697	84,610						
8	,643	6,434	91,044						
9	,479	4,791	95,834						
10	,417	4,166	100,000						

Método de extracción: Análisis de Componentes principales.

ANEXO 6
Análisis de conglomerados de K medias en dos grupos del uso de internet en actividades Académicas

Centros iniciales de los conglomerados

	Conglomerado	
	1	2
¿Aproximadamente cuántas consultas realiza a sus compañeros por mes?	,00	20,00
¿Aproximadamente cuántos videos académicos mira en youtube cada mes?	1,00	16,00
¿Aproximadamente cuántas horas chatea sobre temas académicos cada mes?	,00	5,00
¿Aproximadamente cuántas horas busca información académica en internet cada mes?	,00	60,00

Historial de iteraciones^a

Iteración	Cambio en los centros de los conglomerados	
	1	2
1	12,757	21,381
2	,822	3,664
3	,443	2,064
4	,297	1,377
5	,000	,000

a. Se ha logrado la convergencia debido a que los centros de los conglomerados no presentan ningún cambio o éste es pequeño. El cambio máximo de coordenadas absolutas para cualquier centro es de .000. La iteración actual es 5. La distancia mínima entre los centros iniciales es de 65.192.

Centros de los conglomerados finales

	Conglomerado	
	1	2
¿Aproximadamente cuántas consultas realiza a sus compañeros por mes?	5,72	10,96
¿Aproximadamente cuántos videos académicos mira en youtube cada mes?	3,82	7,81
¿Aproximadamente cuántas horas chatea sobre temas académicos cada mes?	4,82	8,15
¿Aproximadamente cuántas horas busca información académica en internet cada mes?	8,18	35,09

Número de casos en cada conglomerado

Conglomerado	1	325,000
	2	68,000
Válidos		393,000
Perdidos		,000

ANEXO 7
Análisis de conglomerados de K medias en tres grupos del uso de internet en actividades Académicas

Centros iniciales de los conglomerados

	Conglomerado		
	1	2	3
¿Aproximadamente cuántas consultas realiza a sus compañeros por mes?	,00	30,00	3,00
¿Aproximadamente cuántos videos académicos mira en youtube cada mes?	1,00	20,00	2,00
¿Aproximadamente cuántas horas chatea sobre temas académicos cada mes?	,00	4,00	5,00
¿Aproximadamente cuántas horas busca información académica en internet cada mes?	,00	30,00	60,00

Historial de iteraciones^a

Iteración	Cambio en los centros de los conglomerados		
	1	2	3
1	11,404	18,559	15,326
2	1,093	5,014	4,283
3	,844	2,843	1,674
4	,507	1,160	,000
5	,189	,450	,000
6	,104	,251	,000
7	,065	,155	,000
8	,000	,000	,000

a. Se ha logrado la convergencia debido a que los centros de los conglomerados no presentan ningún cambio o éste es pequeño. El cambio máximo de coordenadas absolutas para cualquier centro es de .000. La iteración actual es 8. La distancia mínima entre los centros iniciales es de 44.204.

Centros de los conglomerados finales

	Conglomerado		
	1	2	3
¿Aproximadamente cuántas consultas realiza a sus compañeros por mes?	4,97	9,50	11,60
¿Aproximadamente cuántos videos académicos mira en youtube cada mes?	3,78	5,57	7,68
¿Aproximadamente cuántas horas chatea sobre temas académicos cada mes?	3,94	8,41	7,76
¿Aproximadamente cuántas horas busca información académica en internet cada mes?	5,87	21,58	48,12

Número de casos en cada conglomerado

Conglomerado	1	261,000
	2	107,000
	3	25,000
Válidos		393,000
Perdidos		,000

ANEXO 8
Análisis de conglomerados de K medias en cuatro grupos del uso de internet en actividades Académicas

Centros iniciales de los conglomerados

	Conglomerado			
	1	2	3	4
¿Aproximadamente cuántas consultas realiza a sus compañeros por mes?	5,00	1,00	29,00	7,00
¿Aproximadamente cuántos videos académicos mira en youtube cada mes?	20,00	,00	5,00	5,00
¿Aproximadamente cuántas horas chatea sobre temas académicos cada mes?	,00	,00	20,00	19,00
¿Aproximadamente cuántas horas busca información académica en internet cada mes?	25,00	,00	15,00	60,00

Historial de iteraciones^a

Iteración	Cambio en los centros de los conglomerados			
	1	2	3	4
1	14,341	9,134	14,596	14,656
2	1,789	,709	3,519	2,737
3	,901	,490	1,757	,000
4	,279	,280	1,132	,000
5	,235	,193	,779	,000
6	,000	,113	,396	,000
7	,000	,073	,247	,000
8	,000	,096	,306	,000
9	,000	,135	,399	,000
10	,000	,127	,355	,000
11	,000	,034	,094	,000
12	,000	,047	,129	,000
13	,000	,000	,000	,000

a. Se ha logrado la convergencia debido a que los centros de los conglomerados no presentan ningún cambio o éste es pequeño. El cambio máximo de coordenadas absolutas para cualquier centro es de .000. La iteración actual es 13. La distancia mínima entre los centros iniciales es de 32.265.

Centros de los conglomerados finales

	Conglomerado			
	1	2	3	4
¿Aproximadamente cuántas consultas realiza a sus compañeros por mes?	8,12	3,06	13,11	11,60
¿Aproximadamente cuántos videos académicos mira en youtube cada mes?	5,70	3,56	4,75	7,68
¿Aproximadamente cuántas horas chatea sobre temas académicos cada mes?	8,12	3,40	7,03	7,76
¿Aproximadamente cuántas horas busca información académica en internet cada mes?	23,38	5,32	9,72	48,12

Número de casos en cada conglomerado

Conglomerado	1	86,000
	2	207,000
	3	75,000
	4	25,000
Válidos		393,000
Perdidos		,000

ANEXO 9

Resultados de la clasificación

Número inicial de casos			Grupo de pertenencia pronosticado		Total
			1	2	
Original	Recuento	1	325	0	325
		2	1	67	68
	%	1	100,0	,0	100,0
		2	1,5	98,5	100,0

a. Clasificados correctamente el 99.7% de los casos agrupados originales.

Resultados de la clasificación

Número inicial de casos			Grupo de pertenencia pronosticado			Total
			1	2	3	
Original	Recuento	1	256	5	0	261
		2	3	104	0	107
		3	0	0	25	25
	%	1	98,1	1,9	,0	100,0
		2	2,8	97,2	,0	100,0
		3	,0	,0	100,0	100,0

a. Clasificados correctamente el 98.0% de los casos agrupados originales.

Resultados de la clasificación

Número inicial de casos			Grupo de pertenencia pronosticado				Total
			1	2	3	4	
Original	Recuento	1	86	0	0	0	86
		2	1	203	3	0	207
		3	1	3	71	0	75
		4	0	0	0	25	25
	%	1	100,0	,0	,0	,0	100,0
		2	,5	98,1	1,4	,0	100,0
		3	1,3	4,0	94,7	,0	100,0
		4	,0	,0	,0	100,0	100,0

a. Clasificados correctamente el 98.0% de los casos agrupados originales.

ANEXO 10
Análisis de conglomerados de K medias en dos grupos del uso de internet en actividades de entretenimiento

Centros iniciales de los conglomerados

	Conglomerado	
	1	2
¿ Aproximadamente cuántas horas a la semana chatea por diversión?	,00	28,00
¿Aproximadamente cuántas horas a la semana utiliza redes sociales?	,00	40,00
¿Aproximadamente cuántas horas a la semana descarga música, videos y programas?	6,00	5,00
¿Aproximadamente cuántos videos para entretenimiento mira en youtube cada semana?	,00	30,00

Historial de iteraciones^a

Iteración	Cambio en los centros de los conglomerados	
	1	2
1	14,802	21,273
2	,865	1,831
3	,547	1,197
4	,228	,490
5	,129	,275
6	,000	,000

a. Se ha logrado la convergencia debido a que los centros de los conglomerados no presentan ningún cambio o éste es pequeño. El cambio máximo de coordenadas absolutas para cualquier centro es de .000. La iteración actual es 6. La distancia mínima entre los centros iniciales es de 57.315.

Centros de los conglomerados finales

	Conglomerado	
	1	2
¿Aproximadamente cuántas horas a la semana chatea por diversión?	7,68	20,54
¿Aproximadamente cuántas horas a la semana utiliza redes sociales?	8,23	24,62
¿Aproximadamente cuántas horas a la semana descarga música, videos y programas?	4,18	4,91
¿Aproximadamente cuántos videos para entretenimiento mira en youtube cada semana?	6,85	12,17

Número de casos en cada conglomerado

Conglomerado	1	271,000
	2	122,000
Válidos		393,000
Perdidos		,000

ANEXO 11
Análisis de conglomerados de K medias en tres grupos del uso de internet en actividades de entretenimiento

Centros iniciales de los conglomerados

	Conglomerado		
	1	2	3
¿ Aproximadamente cuántas horas a la semana chatea por diversión?	20,00	24,00	,00
¿Aproximadamente cuántas horas a la semana utiliza redes sociales?	11,00	40,00	,00
¿Aproximadamente cuántas horas a la semana descarga música, videos y programas?	3,00	2,00	6,00
¿Aproximadamente cuántos videos para entretenimiento mira en youtube cada semana?	30,00	2,00	,00

Historial de iteraciones^a

Iteración	Cambio en los centros de los conglomerados		
	1	2	3
1	13,627	13,265	12,076
2	,383	,641	,208
3	,359	,433	,000
4	,736	,172	,259
5	,499	,000	,232
6	,276	,188	,084
7	,322	,000	,153
8	,219	,000	,108
9	,364	,154	,136
10	,795	,983	,000
11	1,350	,676	,514
12	,945	,699	,283
13	,755	,439	,270
14	,272	,141	,168
15	,463	,438	,109
16	,278	,000	,200
17	,265	,000	,198
18	,125	,000	,093
19	,000	,000	,000

a. Se ha logrado la convergencia debido a que los centros de los conglomerados no presentan ningún cambio o éste es pequeño. El cambio máximo de coordenadas absolutas para cualquier centro es de .000. La iteración actual es 19. La distancia mínima entre los centros iniciales es de 37.815.

ANEXO 12
Análisis de conglomerados de K medias en cuatro grupos del uso de internet en actividades de entretenimiento

Centros iniciales de los conglomerados

	Conglomerado			
	1	2	3	4
¿ Aproximadamente cuántas horas a la semana chatea por diversión?	30,00	18,00	2,00	10,00
¿Aproximadamente cuántas horas a la semana utiliza redes sociales?	30,00	11,00	3,00	40,00
¿Aproximadamente cuántas horas a la semana descarga música, videos y programas?	4,00	15,00	2,00	,00
¿Aproximadamente cuántos videos para entretenimiento mira en youtube cada semana?	30,00	2,00	30,00	5,00

Historial de iteraciones^a

Iteración	Cambio en los centros de los conglomerados			
	1	2	3	4
1	10,685	13,795	12,938	10,223
2	4,389	2,136	2,786	2,501
3	1,116	,781	,244	1,796
4	,888	,404	,343	,535
5	,318	,246	,000	,743
6	,252	,171	,000	,419
7	,000	,000	,000	,000

a. Se ha logrado la convergencia debido a que los centros de los conglomerados no presentan ningún cambio o éste es pequeño. El cambio máximo de coordenadas absolutas para cualquier centro es de .000. La iteración actual es 7. La distancia mínima entre los centros iniciales es de 33.749.

Centros de los conglomerados finales

	Conglomerado			
	1	2	3	4
¿ Aproximadamente cuántas horas a la semana chatea por diversión?	24,89	7,28	9,77	16,66
¿Aproximadamente cuántas horas a la semana utiliza redes sociales?	21,07	7,76	9,79	27,69
¿Aproximadamente cuántas horas a la semana descarga música, videos y programas?	5,73	3,96	5,63	3,96
¿Aproximadamente cuántos videos para entretenimiento mira en youtube cada semana?	16,73	4,79	19,00	6,56

Número de casos en cada conglomerado

Conglomerado	1	55,000
	2	222,000
	3	48,000
	4	68,000
Válidos		393,000
Perdidos		,000

ANEXO 13

Resultados de la clasificación

Número inicial de casos			Grupo de pertenencia pronosticado		Total
			1	2	
Original	Recuento	1	268	3	271
		2	5	117	122
	%	1	98,9	1,1	100,0
		2	4,1	95,9	100,0

a. Clasificados correctamente el 98.0% de los casos agrupados originales.

Resultados de la clasificación

Número inicial de casos			Grupo de pertenencia pronosticado			Total
			1	2	3	
Original	Recuento	1	125	1	0	126
		2	6	91	0	97
		3	14	0	156	170
	%	1	99,2	,8	,0	100,0
		2	6,2	93,8	,0	100,0
		3	8,2	,0	91,8	100,0

a. Clasificados correctamente el 94.7% de los casos agrupados originales.

Resultados de la clasificación

Número inicial de casos			Grupo de pertenencia pronosticado				Total
			1	2	3	4	
Original	Recuento	1	54	0	1	0	55
		2	0	217	1	4	222
		3	0	1	47	0	48
		4	0	4	0	64	68
	%	1	98,2	,0	1,8	,0	100,0
		2	,0	97,7	,5	1,8	100,0
		3	,0	2,1	97,9	,0	100,0
		4	,0	5,9	,0	94,1	100,0

a. Clasificados correctamente el 97.2% de los casos agrupados originales.

ANEXO 14

Análisis de conglomerados de K medias en dos grupos del uso de dispositivos electrónicos

Centros iniciales de los conglomerados

	Conglomerado	
	1	2
Smartphone con cámara fotográfica y acceso a internet	10,00	1,00
Teléfono móvil con acceso a internet	10,00	1,00
Computador portátil	10,00	1,00
Tablet (iPad; Galaxy Tab, Kindle, etc)	10,00	1,00

Historial de iteraciones^a

Iteración	Cambio en los centros de los conglomerados	
	1	2
1	3,633	5,674
2	,430	1,085
3	,176	,516
4	,107	,293
5	,098	,257
6	,134	,333
7	,123	,283
8	,041	,092
9	,117	,253
10	,059	,120
11	,066	,134
12	,000	,000

a. Se ha logrado la convergencia debido a que los centros de los conglomerados no presentan ningún cambio o éste es pequeño. El cambio máximo de coordenadas absolutas para cualquier centro es de .000. La iteración actual es 12.

La distancia mínima entre los centros iniciales es de 18.000.

Centros de los conglomerados finales

	Conglomerado	
	1	2
Smartphone con cámara fotográfica y acceso a internet	9,25	5,50
Teléfono móvil con acceso a internet	9,42	6,52
Computador portátil	8,75	5,34
Tablet (iPad; Galaxy Tab, Kindle, etc)	8,13	2,70

Número de casos en cada conglomerado

Conglomerado	1	264,000
	2	129,000
Válidos		393,000
Perdidos		,000

ANEXO 15

Análisis de conglomerados de K medias en tres grupos del uso de dispositivos electrónicos

Centros iniciales de los conglomerados

	Conglomerado		
	1	2	3
Smartphone con cámara fotográfica y acceso a internet	10,00	1,00	7,00
Teléfono móvil con acceso a internet	10,00	1,00	10,00
Computador portátil	1,00	7,00	10,00
Tablet (iPad; Galaxy Tab, Kindle, etc)	1,00	1,00	10,00

Historial de iteraciones^a

Iteración	Cambio en los centros de los conglomerados		
	1	2	3
1	4,627	3,678	2,864
2	1,120	,999	,480
3	,246	,270	,080
4	,054	,105	,000
5	,000	,000	,000

a. Se ha logrado la convergencia debido a que los centros de los conglomerados no presentan ningún cambio o éste es pequeño. El cambio máximo de coordenadas absolutas para cualquier centro es de .000. La iteración actual es 5. La distancia mínima entre los centros iniciales es de 13.077.

Centros de los conglomerados finales

	Conglomerado		
	1	2	3
Smartphone con cámara fotográfica y acceso a internet	8,97	1,88	9,16
Teléfono móvil con acceso a internet	8,51	4,80	9,41
Computador portátil	6,10	5,83	8,84
Tablet (iPad; Galaxy Tab, Kindle, etc)	2,81	2,95	8,95

Número de casos en cada conglomerado

Conglomerado	1	109,000
	2	59,000
	3	225,000
Válidos		393,000
Perdidos		,000

ANEXO 16

Análisis de conglomerados de K medias en cuatro grupos del uso de dispositivos electrónicos

Centros iniciales de los conglomerados

	Conglomerado			
	1	2	3	4
Smartphone con cámara fotográfica y acceso a internet	10,00	1,00	10,00	1,00
Teléfono móvil con acceso a internet	10,00	1,00	1,00	10,00
Computador portátil	1,00	1,00	10,00	10,00
Tablet (iPad; Galaxy Tab, Kindle, etc)	10,00	1,00	3,00	1,00

Historial de iteraciones^a

Iteración	Cambio en los centros de los conglomerados			
	1	2	3	4
1	7,061	4,252	6,160	5,275
2	,877	1,218	,961	1,929
3	,351	,468	1,201	1,093
4	,265	,620	1,272	,532
5	,335	1,010	1,825	,717
6	,124	,357	,734	,284
7	,045	,192	,510	,257
8	,111	,124	,962	,641
9	,080	,124	,515	,228
10	,020	,000	,269	,203
11	,000	,138	,145	,000
12	,000	,000	,190	,187
13	,000	,000	,077	,082
14	,000	,000	,000	,000

a. Se ha logrado la convergencia debido a que los centros de los conglomerados no presentan ningún cambio o éste es pequeño. El cambio máximo de coordenadas absolutas para cualquier centro es de .000. La iteración actual es 14. La distancia mínima entre los centros iniciales es de 12.728.

Centros de los conglomerados finales

	Conglomerado			
	1	2	3	4
Smartphone con cámara fotográfica y acceso a internet	9,16	1,90	8,52	9,44
Teléfono móvil con acceso a internet	9,38	4,86	8,41	8,85
Computador portátil	9,14	5,80	2,97	8,80
Tablet (iPad; Galaxy Tab, Kindle, etc)	9,00	2,78	4,93	2,00

Número de casos en cada conglomerado

Conglomerado	1	212,000
	2	59,000
	3	61,000
	4	61,000
Válidos		393,000
Perdidos		,000

ANEXO 17

Resultados de la clasificación

Número inicial de casos			Grupo de pertenencia pronosticado		Total
			1	2	
Original	Recuento	1	250	14	264
		2	2	127	129
	%	1	94,7	5,3	100,0
		2	1,6	98,4	100,0

a. Clasificados correctamente el 95.9% de los casos agrupados originales.

Resultados de la clasificación

Número inicial de casos			Grupo de pertenencia pronosticado			Total
			1	2	3	
Original	Recuento	1	104	2	3	109
		2	2	57	0	59
		3	3	3	219	225
	%	1	95,4	1,8	2,8	100,0
		2	3,4	96,6	,0	100,0
		3	1,3	1,3	97,3	100,0

a. Clasificados correctamente el 96.7% de los casos agrupados originales.

Resultados de la clasificación

Número inicial de casos			Grupo de pertenencia pronosticado				Total
			1	2	3	4	
Original	Recuento	1	208	4	0	0	212
		2	0	58	1	0	59
		3	0	1	60	0	61
		4	0	0	0	61	61
	%	1	98,1	1,9	,0	,0	100,0
		2	,0	98,3	1,7	,0	100,0
		3	,0	1,6	98,4	,0	100,0
		4	,0	,0	,0	100,0	100,0

a. Clasificados correctamente el 98.5% de los casos agrupados originales.

BIBLIOGRAFÍA

- Abraira, V., & Pérez, A. (1996). Análisis de la varianza. Retrieved March 24, 2015, from http://www.hrc.es/bioest/Anova_1.html
- Aguayo, M. (2012). Cómo hacer una Regresión Logística con SPSS® "paso a paso". Retrieved March 24, 2015, from http://www.fabis.org/html/archivos/docuweb/Regres_log_1r.pdf
- Aguilar-Barojas, S. (2005). Fórmulas para el cálculo de la muestra en investigaciones de salud. *Red de Revistas Científicas de América Latina, El Caribe, España Y Portugal*, 5. Retrieved from <http://drtesis.pe/wp-content/uploads/2014/10/teoria-del-calculodemuestra.pdf>
- Aja Quiroja, L. (2002). Gestión de información, gestión del conocimiento y gestión de la calidad en las organizaciones. Retrieved February 26, 2015, from http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1024-94352002000500004
- Álamo, Ó. (2003). El desafío de la brecha Digital. Retrieved February 26, 2015, from http://ocio.upco.es/sites/ginvinterfacultativo/Biblioteca_de_documentos/1/el_desafio_brecha_digital.pdf
- Albero, M. (2002). Adolescentes e Internet: mitos y realidades de la sociedad de la información. *Zer: Revista de Estudios de Comunicación = Komunikazio Ikasketen Aldizkaria*, 10. Retrieved from <http://www.ehu.eus/zer/hemeroteca/pdfs/zer13-10-albero.pdf>
- Almenara, J. C. (2006). Bases pedagógicas del e-learning. *DIM: Didáctica, Innovación y Multimedia*, (6). Retrieved from <http://www.raco.cat/index.php/DIM/article/view/56479>
- Alzate Molina, C. A., & Gallego Álvarez, G. A. (2012, December 6). Gestión del conocimiento. Seccional Medellín. Retrieved from <http://bibliotecadigital.usbcali.edu.co/jspui/handle/10819/1058>
- Aranda, Y. R., & Sotolongo, A. R. (2013). Integración de los algoritmos de minería de datos 1R, PRISM E ID3 A POSTGRESQL. *Journal of Information Systems and Technology Management*, 10(2), 389–406. doi:10.4301/S1807-17752013000200012
- Ballesteros, A. (2012). Minería de Datos Educativa Aplicada a la Investigación de Patrones de Aprendizaje en Estudiante en Ciencias. *Centro de Investigación En Ciencia Aplicada Y Tecnología Avanzada*.
- Ballesteros, A., Sánchez, D., & García, S. (2013). Minería de datos educativa: Una herramienta para la investigación de patrones de aprendizaje sobre un contexto educativo. Retrieved August 14, 2015, from http://www.lajpe.org/dec13/22-LAJPE_814_bis_Alejandro_Ballesteros.pdf
- Belloch, C. (2010). Entornos Virtuales de Aprendizaje. Retrieved February 25, 2015, from http://moodle2.unid.edu.mx/dts_cursos_md/pos/ED/AV/AM/07/Entornos.pdf
- Bonilla, M. (2001). *Internet y sociedad en América Latina y el Caribe*. FLACSO. Quito. Retrieved from <http://www.flacsoandes.edu.ec/libros/digital/44934.pdf>
- Buckingham, D., & Sefton-Green. (1999). Children, Young People and Digital Technology. *The Journal of Research into New Media Technologies*, 5, 4.
- Cabena, P., Hadjinian, P., Stadler, R., Verhees, J., & Zanasi, A. (1998). *Discovering data mining: from concept to implementation*. Prentice-Hall, Inc.
- Cabero-Almenara, J. (1994). Nuevas tecnologías, comunicación y educación. *Comunicar: Revista científica iberoamericana de comunicación y educación*. Grupo Comunicar. Retrieved from <http://dialnet.unirioja.es/servlet/articulo?codigo=635397&info=resumen&idioma=SPA>
- Carbonell, X., & Beranuy, M. (2007). El adolescente ante las tecnologías de la Información y la comunicación: internet, móvil y videojuegos. Retrieved October 18, 2015, from

- http://www.researchgate.net/profile/Xavier_Carbonell/publication/242199439_EL_ADOL_ESCENTE_ANTE_LAS_TECNOLOGAS_DE_LA_INFORMACION_Y_LA_COMUNICACION_INTERNET_MVIL_Y_VIDEOJUEGOS/links/542173000cf203f155c6bd03.pdf
- Carneiro, R., Toscano, J., & Díaz, T. (2009). Los desafíos de las TIC para el cambio educativo. Retrieved February 26, 2015, from <http://www.oei.es/metas2021/LASTIC2.pdf>
- Caso, J., & Hernández, L. (2007). VARIABLES THAT INFLUENCE ACADEMIC ACHIEVEMENT IN MEXICAN ADOLESCENTS. *Revista Latinoamericana de Psicología*, 39(3), 487–501. Retrieved from http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0120-05342007000300004&lng=en&nrm=iso&tlng=es
- Castaño, C., Fernández, J., & Vásquez, S. . (2008). La E-inclusión y el bienestar social: una perspectiva de género. Retrieved from <http://www.minetur.gob.es/Publicaciones/Publicacionesperiodicas/EconomiaIndustrial/RevistaEconomiaIndustrial/367/139.pdf>
- Castaño-Muñoz, J. (2010, September 13). La desigualdad digital entre los alumnos universitarios de los países desarrollados y su relación con el rendimiento académico. *Universitat Oberta de Catalunya*. Retrieved from <http://openaccess.uoc.edu/webapps/o2/handle/10609/2965>
- Castells, M. (2013). El impacto de internet en la sociedad: una perspectiva global. Retrieved January 7, 2016, from http://teorias.site40.net/gallery/06_impactodeinternet.pdf
- Cerbino, M., & Richero, A. (2007). *Gobernanza, políticas públicas y aplicaciones de Internet*. Quito.
- Coll, C. (2008). Aprender y enseñar con las TIC: expectativas, realidad y potencialidades. Retrieved December 29, 2015, from http://cmapspublic.ihmc.us/rid=1MVHQD5M-NQN5JM-254N/Cesar_Coll_-_aprender_y_ensenar_con_tic.pdf
- De la Fuente, J. (2011). Análisis Factorial.
- Docsetools. (2010). Kendall tau coeficiente de correlación de rangos, Definición, Prueba de hipótesis, Contabilización de las relaciones, Las pruebas de significación, Algoritmos. Retrieved March 24, 2015, from http://docsetools.com/articulos-enciclopedicos/article_93380.html
- Ebner, M., Lienhardt, C., Rohs, M., Meyer, & I. (2010). Microblogs in Higher Education-A chance to facilitate informal and processoriented learning? Retrieved from <http://dx.doi.org/10.1016/>
- ECUADOR, M. D. E. DEL. (2010). Propuesta de estándares educativos. Retrieved from <http://www.educacion.gob.ec/generalidades-pes.html>
- Espinar Ruiz, E., & González Río, M. J. (2008). Jóvenes conectados: las experiencias de los jóvenes con las nuevas tecnologías. Retrieved from <http://rua.ua.es/dspace/handle/10045/33269>
- Espíndola, L., & León, A. (2002). “La Deserción Escolar en América Latina un Tema Prioritario Para la Agenda Regional.” *Revista Iberoamericana de Educación*, 1–17.
- Ferran, P. (2009). Uso autoinformado de Internet en adolescentes: perfil psicológico de un uso elevado de la red. *International Journal of Psychology and Psychological Therapy*, 109–122. Retrieved from <http://www.ijpsy.com/volumen9/num1/225/uso-autoinformado-de-Internet-en-adolescentes-ES.pdf>
- García, M. N. M., Quintales, L. A. M., García, F. J., & Polo, M. (1997). *APLICACIÓN DE TÉCNICAS DE MINERÍA DE DATOS EN LA CONSTRUCCIÓN Y VALIDACIÓN DE MODELOS PREDICTIVOS Y ASOCIATIVOS A PARTIR DE ESPECIFICACIONES DE REQUISITOS DE SOFTWARE*.

- Gargallo, B., Suárez, J. ., & Belloch, M. (2003). La división digital en el proceso de integración de las NTIC en la educación. Diferencias de género entre alumnos de E.S.O. de la comunidad valenciana. Retrieved October 18, 2015, from http://gredos.usal.es/jspui/bitstream/10366/56461/1/TEE2003_V4_divisiondigital.pdf
- González, A., Gisbert, M., Guillen, A., Jiménez, B., LLadó, F., & Rallo, R. (1996). *Las nuevas tecnologías en la educación. En Redes de comunicación, redes de aprendizaje*. Palma: Balears.
- Haberstroh, R. (2008). Data Mining tutorial for Oracle Data Mining.
- 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, M. (2002). “Causas del Fracaso Escolar. In *XIII Congreso de la Sociedad Española de Medicina del Adolescente*.
- Herrera Batista, M. Á. (2009). Disponibilidad, uso y apropiación de las tecnologías por estudiantes universitarios en México: perspectivas para una incorporación innovadora. *Revista Iberoamericana de Educación*. Organización de Estados Iberoamericanos para la Educación, la Ciencia y la Cultura, OEI. Retrieved from <http://dialnet.unirioja.es/servlet/articulo?codigo=2933738&info=resumen&idioma=SPA>
- IBM. (2013, January 1). Análisis de clúster Jerárquico. Retrieved September 22, 2015, from http://www-01.ibm.com/support/knowledgecenter/SSLVMB_22.0.0/com.ibm.spss.statistics.help/sps/base/idh_clus.htm?lang=es
- Jiménez, J. (2010). El papel de las TIC en el desarrollo: una mirada desde la construcción social de la tecnología en el caso Ecuatoriano. *Biblat*. Retrieved from <http://132.248.9.34/hevila/iconosRevistadecienciassociales/2010/no37/7.pdf>
- Jiménez, N. (2008). *Los medios de comunicación frente a la revolución de la información*. Retrieved from <http://www.rppnet.com.ar/mediatico.htm>
- Jolliffe, A., Ritter, J., & Stevens, D. (2001). *The Online Learning Handbook: Developing and Using Web-based Learning*. Retrieved from <https://books.google.com/books?hl=es&lr=&id=2wi7uPGt0c4C&pgis=1>
- Jonassen, D., Howland, J., Moore, J., & Marra, R. (2002). *Learning to Solve Problems with Technology: A Constructivist Perspective (2nd Edition)*. {Prentice Hall}.
- Larrañaga, P., & Inza, I. (2006). Introducción a la minería de datos.
- Luckin, R. (2010). Re-designing Learning Contexts. Retrieved December 27, 2015, from <http://tel.ioe.ac.uk/wp-content/uploads/2010/04/re-designinglearningcontextsdiscounflyer.pdf>
- Manzano, J. (2008). Análisis de Conglomerados. Retrieved March 23, 2015, from <http://www.uv.es/aldas/resources/Docencia/UNR/CONGLOMERADOS.pdf>
- Molano, E. G., Sánchez, O. N., & Castillo, J. N. P. (2013, August 2). APRENDIZAJE CON MODELO TCPK Y E-LEARNING DE LA LÓGICA DE PROGRAMACIÓN. *Vínculos*. Retrieved from <http://revistas.udistrital.edu.co/ojs/index.php/vinculos/article/view/4287/8779>
- Moreno, M., & Martínez, R. (2011). Modelo Clasificador para Predecir el Desempeño Escolar Terminal de n Estudiante. Retrieved August 17, 2015, from http://www.iiis.org/CDs2011/CD2011CSC/CISCI_2011/PapersPdf/CA499XF.pdf
- Naval, C. (Concepción), Sádaba, C. (Charo), & Bringué, X. (Xavier). (2003). *Impacto de las Tecnologías de la Información y de la Comunicación (TIC) en las Relaciones Sociales de los Jóvenes Navarros*. Gobierno de Navarra. Instituto Navarro de Deporte y Juventud.

- Retrieved from <http://dadun.unav.edu/handle/10171/18444>
- Ortega, E., & Guell, P. (2006). Las Nuevas Tecnologías, ¿un salto al futuro? Retrieved February 26, 2015, from <http://chanchorangutan.cl/idh/informes/2006-las-nuevas-tecnologias-un-salto-al-futuro/>
- Pérez, A. (2006). *Un modelo para sistemas complejos basado en técnicas de elementos finitos bidimensionales*. Universidad de Alicante. Retrieved from file:///C:/Users/Wilson/Downloads/tesis_doctoral_antonio_perez.pdf
- Pluma, N. (2014). Procedimiento de Tablas de Contingencia. Retrieved December 16, 2015, from <http://nayzul.blogspot.com/2014/03/procedimiento-tablas-de-contingencia.html>
- Quevedo Ricardi, F. (2011). The chi-square. *Medwave*, 11(12), e5266–e5266. doi:10.5867/medwave.2011.12.5266
- Ramesh, V. (2011). Performance Analysis of Data Mining Techniques for Placement Chance Prediction. Retrieved August 14, 2015, from <http://www.ijser.org/researchpaper%5cperformance-analysis-of-data-mining-techniques-for-placement-chance-prediction.pdf>
- Riquelme, J. C., Ruiz, R., & Gilbert, K. (2006). Minería de datos: conceptos y tendencias. *Revista Iberoamericana de Inteligencia Artificial*, 11–18.
- Rodríguez, D. (2006). Modelos para la creación y gestión del conocimiento : una aproximación teórica, (37), 25–39. Retrieved from <http://www.raco.cat/index.php/Educuar/article/view/58019>
- Romero, C., Ventura, S., & Hervás, C. (2005). Estado actual de la aplicación de la minería de datos a los sistemas de enseñanza basada en web. Retrieved August 4, 2015, from <http://www.lsi.us.es/redmidas/CEDI/papers/189.pdf>
- Romero, C., Ventura, S., Pechenizkiy, M., & Baker, R. (2010). *Handbook of Educational Data Mining*. CRC Press. Retrieved from https://books.google.com.ec/books?hl=es&lr=&id=u5aWVw0uQJMC&oi=fnd&pg=PP1&dq=Handbook+of+Educational+Data+Mining&ots=KB1Wly_47G&sig=cDgNcp90E_riJ2RihssOsw1buQ#v=onepage&q=Handbook+of+Educational+Data+Mining&f=false
- Rosario, J. (2005). La Tecnología de la Información y la Comunicación (TIC). Su uso como Herramienta para el Fortalecimiento y el Desarrollo de la Educación Virtual- Archivo de artículos del Observatorio para la CiberSociedad. Retrieved February 26, 2015, from <http://www.cibersociedad.net/archivo/articulo.php?art=218>
- Ruiz, F., Díaz, A., Chang, H., & Echeverría, F. (2009). Sistema de predicción y recomendación personalizada basada en ranqueo de ítems homogéneos usando filtrado colaborativo. Retrieved from <https://www.dspace.espol.edu.ec/bitstream/123456789/771/1/1370.pdf>
- Sánchez Galvis, M. (2010). Implicaciones de Género en la Sociedad de la Información: Un Análisis desde los Determinantes de Uso de Internet en Chile y México. *Journal of Technology Management & Innovation*, 5(1), 108–126. doi:10.4067/S0718-27242010000100009
- Santillan, M. (2006). Tecnologías de la información y de la comunicación en la educación. Retrieved January 6, 2016, from <http://www.comie.org.mx/documentos/rmie/v11/n28/pdf/rmie/v11n28scB01n01es.pdf>
- Serrano, F., & Sánchez, P. (2012). ANÁLISIS CUANTITATIVO DE DATOS EN CIENCIAS SOCIALES CON EL SPSS (I) Tablas de contingencia y pruebas de asociación. Retrieved from http://digitum.um.es/jspui/bitstream/10201/27921/1/SPSS_TCONTINGENCIA.pdf
- Terrádez, M., & Juan, Á. (2012). Análisis de varianza (ANOVA). Retrieved March 24, 2015, from <http://www.uoc.edu/in3/emath/docs/ANOVA.pdf>
- Thomas, D., & Brown, J. S. (2011). *A New Culture of Learning: Cultivating the Imagination for*

- a *World of Constant Change*. CreateSpace. Retrieved from <https://books.google.com/books?id=p1tBYgEACAAJ&pgis=1>
- Torrado, F., & Berlanga, V. (2012). Análisis Discriminante mediante SPSS. Retrieved March 8, 2015, from <http://diposit.ub.edu/dspace/bitstream/2445/45344/1/627683.pdf>
- Torrado, M. (2011). Taller Minería de datos aplicados a la educación, 0.
- Torres Díaz, J. C., & Infante Moro, A. (2011). Digital Divide in Universities: Internet Use in Ecuadorian Universities. *Comunicar*, 19(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 en un grupo de universidades ecuatorianas presenciales*. Universitat Oberta de Catalunya.
- Twining, P. (2002). Conceptualising Computer Use in Education: Introducing the Computer Practice Framework (CPF). *British Educational Research Journal*, 28(1), 95–110. doi:10.1080/01411920120109775
- UB. (2010). MEDIDAS DE ASOCIACIÓN PARA DATOS ORDINALES. Retrieved March 24, 2015, from http://www.ub.edu/aplica_infor/spss/cap3-5.htm
- UCM. (2011a). Analisis de conglomerados: El procedimiento conglomerados K medias. Retrieved March 23, 2015, from http://pendientedemigracion.ucm.es/info/socivmyt/paginas/D_departamento/materiales/analisis_datosyMultivariable/21conglk_SPSS.pdf
- UCM. (2011b). Análisis de correlación lineal: Los procedimientos, correlaciones bivariadas y correlaciones parciales. Retrieved March 24, 2015, from http://pendientedemigracion.ucm.es/info/socivmyt/paginas/D_departamento/materiales/analisis_datosyMultivariable/17corlin_SPSS.pdf
- UV. (2011). Coeficiente de Pearson. Retrieved March 24, 2015, from http://www.uv.es/webgid/Descriptiva/31_coeficiente_de_pearson.html
- Vera, C., Morales, C., & Soto, S. (2012). Predicción del fracaso Escolar mediante Técnicas de Minería de Datos. *Revista Iberoamericana de Tecnologías Del Aprendizaje*, 109. Retrieved from <http://rita.det.uvigo.es/201208/uploads/IEEE-RITA.2012.V7.N3.A1.pdf>
- Vieira, L., Ortiz, L., & Carvajal, S. (2009). *Introducción a la Minería de Datos*. Rio de Janeiro: E-papers servicios editoriales.
- Vivanco, M. (2005). *Muestreo Estadístico. Diseño Y Aplicaciones*. Retrieved from https://books.google.com/books?hl=es&lr=&id=-_gr5l3LbplC&pgis=1
- Weiss, S., & Indurkha, N. (1998). *La minería de datos predictivo: una guía práctica*. Morgan Kaufmann.
- Zelwyn, N. (2013). Internet y Educación. Retrieved from <http://chchsws.netii.net/gallery/09/intenetyeducacion.pdf>