



UNIVERSIDAD TÉCNICA PARTICULAR DE LOJA
La Universidad Católica de Loja

ÁREA TÉCNICA

TÍTULO DE INGENIERO EN INFORMÁTICA

La Educación Virtual en el Ecuador
“Análisis de los usos de la tecnología en los estudiantes de la Universidad Nacional de Loja”

TRABAJO DE TITULACIÓN.

AUTOR: Jaramillo Tapia, Tatiana Elizabeth
DIRECTOR: Jara Roa, Dunia Inés

CENTRO UNIVERSITARIO LOJA

2015

APROBACIÓN DEL DIRECTOR DEL TRABAJO DE TITULACIÓN

Magister.

Dunia Inés Jara Roa.

DIRECTORA DEL TRABAJO DE TITULACIÓN.

De mi consideración:

El presente trabajo de titulación: Análisis de los usos de la tecnología en los estudiantes de la Universidad Nacional de Loja realizado por Jaramillo Tapia Tatiana Elizabeth, ha sido orientado y revisado durante su ejecución, por cuanto se aprueba la presentación del mismo.

Loja, Noviembre del 2015

f).....

DECLARACIÓN DE AUTORÍA Y CESIÓN DE DERECHOS

“Yo Tatiana Elizabeth Jaramillo Tapia declaro ser autor (a) del presente trabajo de titulación: Análisis de los Usos de la Tecnología en los Estudiantes de la Universidad Nacional de Loja, de la Titulación de Informática, siendo la Mgs. Dunia Inés Jara Roa 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 que se realicen a través, o con el apoyo financiero, académico o institucional (operativo) de la Universidad”.

F.....
Jaramillo Tapia Tatiana Elizabeth
1104101314

DEDICATORIA

La presente tesis está dedicada con mucho amor y cariño, primeramente a Dios por haberme brindado salud y permitirme llegar a este momento muy especial en mi vida.

A mis padres Rodrigo y Lidia por su amor, trabajo y sacrificio en todos estos años he logrado convertirme en lo que ahora soy, sembrando en mi buenos valores y virtudes que me han ayudado en momentos difíciles, en especial a mi madre quien me ayudó con mucho amor en las largas horas y días del cuidado de mis pequeñas.

A mí querido esposo José Vera que con su amor, apoyo y comprensión me ha mantenido en pie de lucha durante todo este tiempo.

En especial a mis dos hermosas hijas: Angelita y Karito por ser mi mayor fuente de inspiración para culminar con éxito este proyecto, por soportar a temprana edad mi ausencia, y las largas horas frente a un computador, pero a pesar de ello, cada minuto que teníamos de tiempo lo aprovechábamos, en lo que una sola sonrisa o patadita me llenaba de fuerza y ánimos.

A toda mi familia en general, por sus palabras alentadoras y su amor que han contribuido para el logro de mis objetivos.

AGRADECIMIENTO

Quiero expresar en estas cortas palabras mis más sinceros agradecimientos del presente trabajo investigativo a la Universidad Técnica Particular de Loja, por darme la oportunidad de forjarme como su estudiante y ser ahora un profesional.

Quiero agradecer a cada una de las personas que han sabido darme su apoyo incondicional en el trascurso de este objetivo tan anhelado, de manera muy especial a mis directores de tesis: la Mgs. Inés Jara, al Mgs. Juan Carlos Torres y a la secretaria de la modalidad abierta y a distancia de la carrera de informática, la Lic. Lidia Villacís por su paciencia, motivación y confianza depositada en mi persona y por brindarme la oportunidad de involucrarme en el desarrollo de este proyecto investigativo, ayudándome a adquirir nuevos conocimientos que me ayudaran durante mi carrera profesional.

Un especial agradecimiento a dios, por haberme otorgado una hermosa familia quienes han creído en mí, brindándome ejemplos de superación de humildad y sacrificio para que pueda valorar todo lo que tengo hasta el día de hoy. A mis padres, hermanos, mi esposo e hijas quienes supieron entenderme en mis momentos buenos y malos.

A todos mi gratitud infinita.

Tatiana

ÍNDICE DE CONTENIDOS

APROBACIÓN DEL DIRECTOR DEL TRABAJO DE FIN DE TITULACIÓN	ii
DECLARACIÓN DE AUTORÍA Y CESIÓN DE DERECHOS.....	iii
DEDICATORIA	iv
AGRADECIMIENTO	v
ÍNDICE DE CONTENIDOS.....	vi
ÍNDICE DE TABLAS.....	ix
RESUMEN.....	1
ABSTRACT	2
CAPÍTULO 1: INTRODUCCIÓN	3
1.1. Introducción.....	4
1.2. Objetivos.	5
1.3. Preguntas de Investigación.....	6
1.4. Hipótesis.....	6
1.5. Estructura del documento.....	7
CAPÍTULO 2: MARCO TEÓRICO.....	8
2.1. El uso del Internet.....	9
2.2. Brecha digital: definiciones generales.....	11
2.2.1. Aspectos de la brecha digital.....	13
2.2.2. Causas o factores que influyen en la brecha digital.....	15
2.2.3. Desigualdad digital en América Latina y el Ecuador.....	22
2.3. El rendimiento académico y el uso del Internet.....	24
2.3.1. El uso de las redes sociales y su relación con el aprendizaje.....	27
2.3.2. El uso de las redes sociales y su incidencia en el rendimiento académico	31
2.4. Minería de datos.....	33
2.4.1. Tipos de datos.....	34
2.4.2. Tipos de modelo.....	35
2.4.3. Proceso de extracción de conocimiento.....	37
2.4.4. Técnicas de minería de datos.....	39
2.4.5. Reconocimiento de patrones	42
CAPÍTULO 3: METODOLOGÍA	54
3.1. Población y tamaño de la muestra.....	55
3.2. Análisis e interpretación de los datos.....	56
3.2.1. Fase de integración y recopilación de datos.....	57

3.2.2.	Fase de selección, limpieza y transformación de datos.	58
3.2.3.	Fase de minería de datos.	60
3.2.3.	Fase de evaluación e interpretación de los datos.	65
CAPÍTULO 4:	RESULTADOS.....	69
4.1.	Fase de Minería de datos.	70
4.1.1.	Datos generales de la población universitaria.....	70
4.1.2.	Uso de Internet en las preferencias del estudiante.	71
4.1.3.	Uso del Internet en las actividades académicas.	76
4.1.4.	Uso del Internet para entretenimiento.....	81
4.1.5.	Generalidades sobre el rendimiento académico.	85
4.1.6.	Reconocimiento de patrones.	86
4.2.	Fase de Evaluación e Interpretación de los datos.....	95
4.2.1.	Incidencia del nivel de ingresos en el uso del Internet para el aprendizaje.	95
4.2.2.	Incidencia del nivel de ingresos con el uso de Internet para entretenimiento.....	97
4.2.3.	Incidencia del uso de la tecnología en el aprendizaje sobre el rendimiento académico.....	98
4.2.4.	Incidencia del uso de la tecnología para entretenimiento en el rendimiento académico.....	99
CAPÍTULO 5:	DISCUSIÓN DE RESULTADOS	101
5.1.	Internet en las preferencias de los estudiantes, en las actividades académicas y para entretenimiento.....	102
5.2.	Efecto del nivel de ingresos sobre los usos de Internet en actividades académicas.	105
5.3.	Efecto del nivel de ingresos sobre los usos de Internet para el entretenimiento.	106
5.4.	Efecto de los usos de Internet en actividades académicas sobre el rendimiento académico.....	107
5.5.	Efecto de los usos de Internet en entretenimiento sobre el rendimiento académico	109
CAPÍTULO 6:	CONCLUSIONES Y RECOMENDACIONES	111
6.1.	Conclusiones.....	112
6.2.	Recomendaciones.....	113
REFERENCIAS	115
ANEXOS.....		124
Anexo 1:	Población universitaria del periodo académico Octubre 2014- Febrero 2015.....	125
Anexo 2:	Encuesta para estudiantes	126
Anexo 3:	Carta dirigida al rector de la Universidad en estudio.....	128
Anexo 4:	Contestación favorable de la carta dirigida al rector de la Universidad en estudio.....	129
Anexo 5:	Análisis Factorial para las actividades académicas.....	130
Anexo 6:	Análisis Factorial para entretenimiento	131
Anexo 7:	Análisis clúster para clasificar en dos grupos según el uso del Internet en	

actividades académicas.....	132
Anexo 8: Análisis clúster para clasificar en tres grupos según el uso del Internet en actividades académicas.....	133
Anexo 9: Análisis clúster para clasificar en cuatro grupos según el uso del Internet en actividades académicas.....	134
Anexo 10: Análisis clúster para clasificar en dos grupos según los usos de la tecnología para el entretenimiento.	135
Anexo 11: Análisis clúster para clasificar en tres grupos según los usos de la tecnología para el entretenimiento.	136
Anexo 12: Análisis clúster para clasificar en cuatro grupos según los usos de la tecnología para el entretenimiento.	137
Anexo 13: Análisis clúster para clasificar en dos grupos según el nivel de uso de los dispositivos.....	138
Anexo 14: Análisis clúster para clasificar en tres grupos según el nivel de uso de los dispositivos.....	139
Anexo 15: Análisis clúster para clasificar en cuatro grupos según el nivel de uso de los dispositivos.....	140
Anexo 16: Ajuste del modelo del uso de Internet en actividades académicas.....	141
Anexo 17: Ajuste del modelo en función de los usos de tecnología en el entretenimiento ...	142
Anexo 18: Ajuste del modelo del uso de la tecnología en el aprendizaje sobre el rendimiento académico.....	143
Anexo 19: Ajuste del modelo del uso de la tecnología para entretenimiento sobre el rendimiento académico.....	144
Anexo 20: Información General del proyecto “Educación virtual en Ecuador”	145

ÍNDICE DE TABLAS

Tabla 1. Posiciones de países latino-americanos dentro del Networked Readness Index 2014	23
Tabla 2. Relación entre las técnicas con las tareas de minería de datos	40
Tabla 3. Variables inmersas en la transformación de los datos	59
Tabla 4. Variables seleccionadas para la comprobación de hipótesis	60
Tabla 5. Discriminación para actividades académicas de 3 grupos	64
Tabla 6. Discriminación para entretenimiento de 3 grupos	65
Tabla 7. Valores del Chi-cuadrado de la H1 y H2	66
Tabla 8. Valores del Chi-cuadrado de la H3 y H4	67
Tabla 9. Días de conexión a Internet	72
Tabla 10. Nivel de conocimiento en uso del Internet	73
Tabla 11. Horas que se conecta al día	74
Tabla 12. Años de experiencia en Internet	74
Tabla 13. Nivel de uso de los dispositivos móviles	76
Tabla 14. Consultas realizadas a los profesores por mes	77
Tabla 15. Consultas realizadas a los compañeros por mes	78
Tabla 16. Descargas de recursos educativos por mes	78
Tabla 17. Videos académicos vistos en youtube cada mes	79
Tabla 18. Búsqueda de información académica en Internet al mes	81
Tabla 19. Ingreso a la biblioteca virtual	81
Tabla 20. Horas de chat semanal por diversión	82
Tabla 21. Horas semanales empleadas en redes sociales	82
Tabla 22. Horas semanales dedicadas a los juegos en línea	83
Tabla 23. Videos para entretenimiento vistos en youtube a la semana	84
Tabla 24. Porcentaje de precisión de la clasificación (académica)	88
Tabla 25. Nivel de discriminación para actividades académicas	88
Tabla 26. Porcentaje de precisión de la clasificación (entretenimiento)	91
Tabla 27. Nivel de discriminación para entretenimiento	91
Tabla 28. Porcentaje de precisión de la clasificación de dispositivos	94
Tabla 29. Coeficientes del modelo de regresión logística (H1)	96
Tabla 30. Coeficientes del modelo de regresión logística (H2)	98
Tabla 31. Coeficientes del modelo de regresión logística binaria (H3)	99
Tabla 32. Coeficientes del modelo de regresión logística binaria (H4)	100

Tabla 33. Análisis de los datos de la H1	106
Tabla 34. Análisis de los datos de la H2	107
Tabla 35. Análisis de los datos de la H3	107
Tabla 36. Análisis de los datos de la H4	109

ÍNDICE DE FIGURAS

Figura 1. Acceso provincial a Internet en el Ecuador, hasta Diciembre 2014	11
Figura 2. Índice Networked Readiness 2014.....	13
Figura 3. Uso del Internet en el Ecuador por quintiles de ingreso 2008-2010	20
Figura 4. Analfabetismo digital por género en el Ecuador, según el censo 2010.....	21
Figura 5. Analfabetismo digital por auto-identificación étnica en el Ecuador, hasta el año 2010.	21
Figura 6. Proceso de KDD	37
Figura 7. Fases del KDD.....	56
Figura 8. Distribución de estudiantes por género.....	70
Figura 9. Distribución de estudiantes por edad.....	70
Figura 10. Distribución de estudiantes por los Ingresos familiares	71
Figura 11. Conexión habitual a Internet	72
Figura 12. Estudiantes que tienen blog, cuenta en youtube y en del.icio.us.....	75
Figura 13. Ingreso a la plataforma virtual	77
Figura 14. Participación de foros cada mes.....	79
Figura 15. Participación de post o tweets cada mes.....	80
Figura 16. Chat destinado a temas académicos por mes.....	80
Figura 17. Horas de descargas de música, videos y programas.....	83
Figura 18. Seguidores en twitter, facebook y linkedIn.....	84
Figura 19. Módulo aprobado/no aprobado (rendimiento académico).....	85
Figura 20. Perfil del uso del Internet en el aspecto académico	89
Figura 21. Perfil del uso de Internet para entretenimiento.....	92
Figura 22. Perfil del uso de dispositivos	94

RESUMEN

La investigación determina la incidencia de los ingresos económicos familiares de los estudiantes de la Universidad Nacional de Loja sobre el uso del Internet en las actividades académicas y de entretenimiento; así mismo, busca comprobar la incidencia del uso del Internet en relación al rendimiento académico.

Para ello se utilizó el proceso de extracción de conocimiento (KDD) conjuntamente con las tareas y algoritmos como la clusterización, para la agrupación de los estudiantes acorde a sus características comunes, resultando los clústers denominados perfiles: perfil de uso de Internet en actividades académicas y el perfil del uso del Internet para entretenimiento, cada uno conformado de tres niveles. Los resultados obtenidos se los verificaron con el análisis discriminante con el objetivo de obtener el porcentaje de exactitud de los grupos que mejor discriminen y finalmente para la verificación de las hipótesis planteadas se aplicó la técnica de minería de datos predictiva como la regresión logística multinomial y binomial.

PALABRAS CLAVES: brecha digital, uso del Internet, rendimiento académico, análisis clúster, análisis discriminante, regresión logística, minería de datos.

ABSTRACT

The investigation determined the incidence of family income of students from the National University of Loja on using the Internet in academic activities and entertainment; It also seeks to check the impact of the use of the Internet in relation to academic performance.

For this, the process of extracting knowledge (KDD) was used in conjunction with tasks and clustering algorithms such as, for grouping students according to their common characteristics, resulting clusters called profiles: profile of Internet use in academic activities and profile of Internet use for entertainment, each consisting of three levels. The results obtained are verified with the discriminant analysis in order to obtain the percentage of accuracy of better discriminate groups and finally for verifying the hypotheses technique predictive data mining applied as the binomial and multinomial logistic regression.

KEYWORDS: digital divide, Internet use, academic performance, cluster analysis, discriminant analysis, logistic regression, data mining.

CAPÍTULO 1: INTRODUCCIÓN

1.1. Introducción.

Se ha visto grandes avances en los países en vías de desarrollo como también en los países desarrollados en cuanto a la era tecnológica y la comunicación, en diversos campos en especial el educativo, donde el uso de la tecnología se expande aceleradamente con gran facilidad y rapidez, incluyendo una gran cantidad de información que se encuentra almacenada en la web (Tello, Sosa, Castillo, y Flores, 2015).

La educación afronta algunos retos y desafíos en estos tiempos, en lo que se refiere a la eficiencia que debe tener el proceso de enseñanza y aprendizaje, siendo este dinámico y flexible, tanto para docentes como para estudiantes (Salinas, 2013). Los estudiantes que hacen uso de las herramientas tecnológicas se involucran en una experiencia productiva a nivel social, académico y de entretenimiento, ya que se ven inmersos a comunicarse con sus compañeros de clase, con sus profesores, amigos o conocidos, permitiéndoles en su labor académica tener debates sobre los temas de estudio en forma grupal o individual; así mismo, el uso del Internet favorece a los profesores ayudándoles a actualizar sus materiales didácticos, fomentando nuevos procedimientos pedagógicos y metodológicos (Meso, Pérez, y Mendiguren, 2011), como colocar temas de discusión haciendo uso de herramientas tecnológicas, como por ejemplo, en una plataforma virtual, en donde los temas propuestos vayan acorde con los objetivos de la materia que están revisando en el periodo académico; esto conlleva a mantener el desarrollo de sus clases de forma más consistente y obteniendo por parte de los estudiantes mejores conocimientos y versatilidad en su aprendizaje; por tal razón, el uso de la tecnología favorece el intercambio de información entre los alumnos y los docentes de forma más dinámica a través de la web (Meso, Pérez, y Mendiguren, 2011). Sin embargo, las herramientas tecnológicas presentan sus ventajas y desventajas, dependiendo del buen o mal uso que se le proporcione, los estudiantes que incorporan las herramientas tecnológicas con un uso ético adecuado, tendrán un mejor desarrollo de las actividades de aprendizaje mejorando así su rendimiento académico.

Cabe indicar, que existen otros factores involucrados en el uso de las herramientas tecnológicas, lo que causa diferentes desigualdades sociales, tanto de acceso como de uso (Torres, 2012). Es necesario conocer los usos tecnológicos que benefician el rendimiento académico así como sus determinantes para disminuir la desigualdad entre el uso del Internet y el rendimiento académico.

El presente trabajo investigativo es considerado positivo e interesante por abordar un tema de importancia y sobre todo por estar en auge en estos tiempos, con él se realiza un levantamiento de información sobre el uso de la tecnología en el sistema educativo de la universidad Nacional de Loja (UNL), analizando como los estudiantes utilizan la tecnología en sus actividades académicas y de entretenimiento. Para conocer la realidad se hace el levantamiento de la información y se aplica las técnicas de minería de datos conjuntamente con las herramientas estadísticas para verificar la incidencia de los ingresos económicos familiares sobre las actividades académicas y de entretenimiento, como también comprobar la incidencia del uso del Internet sobre el rendimiento académico. Se busca agrupar los perfiles de los estudiantes en el uso de Internet, categorizándolos de acuerdo a sus características más comunes. El resultado de estas clasificaciones dieron 3 perfiles: perfil del uso del Internet en actividades académicas, perfil del uso del Internet para entretenimiento y perfil del nivel de uso de dispositivos. El primer perfil cuenta con tres niveles: pasivo, medio e interactivo; el segundo perfil consta de tres niveles: inactivo, intermedio, dinámico; y finalmente el tercer perfil está compuesto de dos niveles: medio y alto. Además, se establecen relaciones entre los niveles de ingresos familiares y el rendimiento académico, sobre el aprendizaje y las actividades de entretenimiento y diversión, analizando el nivel de incidencia que existe entre estas variables en estudio.

En el ámbito de la minería de datos para la creación del modelo, la metodología utilizada fue el proceso de extracción de conocimiento (KDD), la misma que es una de las más empleadas en la generación de proyectos de minería de datos, en la cual se utilizó las técnicas de la clusterización, la discriminación y la regresión logística.

1.2. Objetivos.

Los objetivos enfocados para este tema de estudio están detallados en base a los objetivos generales del proyecto (Anexo 20).

Objetivo General.

- Identificar la incidencia del nivel de ingresos y rendimiento académico sobre las actividades académicas y de entretenimiento de los estudiantes de la UNL.

Objetivos Específicos.

- Levantar información de la Educación virtual en la UNL con una línea de base respecto al uso de la tecnología en el sistema educativo del país.
- Categorizar a los estudiantes en grupos homogéneos
- Identificar el nivel de incidencia de los ingresos familiares sobre los perfiles de usuarios.
- Buscar la existencia de relaciones entre los usos de la tecnología y el éxito académico que alcanzan los estudiantes lo que permitirá determinar el real efecto de las herramientas tecnológicas dentro del sistema educativo.

1.3. Preguntas de Investigación.

Las preguntas de investigación planteadas son:

1. ¿Cómo se relacionan los niveles de ingreso de las familias de los estudiantes universitarios con los usos de Internet en actividades académicas y de entretenimiento?
2. ¿Cómo se relacionan el rendimiento académico y los usos de Internet en actividades académicas y de entretenimiento?

1.4. Hipótesis.

Hipótesis relacionadas a primera pregunta de investigación

- **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 relacionadas a la segunda pregunta de investigación.

- **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.

1.5. Estructura del documento.

El presente trabajo investigativo está conformado por seis capítulos, donde se han desarrollados algunos temas fundamentales y generales en torno al “Análisis de los usos de la tecnología en los estudiantes de la UNL”, cuyos temas consisten principalmente en el uso del Internet, brecha digital, el rendimiento académico y minería de datos.

En el capítulo I, se describe la introducción, los objetivos, las preguntas de investigación y las hipótesis planteadas. El capítulo II, presenta el Marco Teórico sobre los principales temas que aborda la investigación como el uso del Internet, brecha digital, el rendimiento académico y minería de datos. El capítulo III, se refiere a la metodología empleada para alcanzar los objetivos deseados en el presente estudio, describiendo la población y el tamaño de la muestra, el proceso KDD con sus respectivas fases: fase de integración y recopilación de los datos, fase de selección, limpieza y transformación de los datos, principalmente la fase de minería de datos y la fase de evaluación e interpretación de los datos. El capítulo IV, se presenta los resultados alcanzados durante la investigación, detallando los datos generales de la población universitaria, el uso del Internet: sobre las preferencias de los estudiantes, en las actividades académicas y para el entretenimiento; además, se describe las generalidades sobre el rendimiento académico; a continuación, se determina el proceso del reconocimiento de patrones que permitan establecer las características de los estudiantes denominándolas: perfiles de los estudiantes por medio de la técnica de minería de datos como es la clusterización; y finalmente se muestra la verificación de las hipótesis con la técnica de regresión logística a través de la fase de evaluación e interpretación de los datos concerniente al proceso KDD. En el capítulo V, se presenta la discusión de resultados; en donde se puntualiza el efecto que causa los ingresos en las actividades académicas como de entretenimiento; también, se describe la incidencia de los usos de la tecnología (académicas/entretenimiento) sobre el rendimiento académico, determinando las hipótesis que se han cumplido durante el presente estudio. Finalmente, en el Capítulo VI, se extraen las conclusiones y recomendaciones logradas en el proceso investigativo.

CAPÍTULO 2: MARCO TEÓRICO

2.1. El uso del Internet.

Como es debido en toda investigación se iniciará con un compendio de los conceptos relacionados y necesarios para llevar a cabo el trabajo, en este apartado se definirá y se aportará todos los conocimientos científicos relacionados al uso de la Internet.

Para entender mejor el significado que engloba a la palabra Internet se tomará el concepto otorgado por Reno (1996, citado por Álvarez , 2011) que dice:

Internet no es una entidad física o tangible, sino más bien es una red gigante de computadoras interconectadas, es una red de redes. Esta gran red está formada por una colección de miles de redes enlazadas a través de una serie de protocolos técnicos comunes que hacen posible que los usuarios de cualquiera de esas redes se comuniquen o usen los servicios de cualquiera de las demás redes (p.1).

Si bien es cierto, el concepto está muy bien definido pero puede que tal vez sea algo confuso para aquellas personas que no estén muy familiarizadas con términos informáticos, es por ello que Álvarez (2011) resume el significado de la siguiente forma: “la Internet es un medio para difundir información, para el debate público y la comunicación personal, para el comercio y la prestación de servicios; sin importar la ubicación geográfica” (p. 2). De esta definición se puede colegir que Internet es una herramienta fundamentalmente comunicativa tanto para particulares, como para empresas y estados.

De igual forma sería conveniente poner en evidencia un poco de la evolución que ha tenido la Internet durante los últimos años, cabe destacar que no se busca realizar una reseña histórica de la red, sino más bien una breve descripción de cómo ha evolucionado la red y en que ha consistido cada una de sus etapas, para ello se tomará la descripción que ofrece Belloch (2012):

- **Web 1.0:** se basa en la sociedad de la información, en medios de entretenimiento y consumo pasivo (medios tradicionales, radio, TV, email). Las pinas web son estáticas y con poca interacción con el usuario, en esta etapa Belloch no indica que las pinas web son solo para leer.
- **Web 2.0:** se basa en la sociedad del conocimiento, la autogeneración de contenido, en medios de entretenimiento y consumo activo. Belloch indica que en esta etapa las pinas web se caracterizan por ser dinámicas e interactivas en donde el usuario comparte información y

recursos con otros usuarios. En esta etapa se han creado redes sociales que conforman comunidades en donde los usuarios pueden incluir sus opiniones, fotografías, y comunicarse con el resto de miembros de su comunidad.

- **Web 3.0:** las innovaciones que se están produciendo en estos momentos se basan en sociedades virtuales, realidad virtual, web semántica, búsqueda inteligente (p. 5).

Se evidencia que la conectividad entre personas y empresas avanza a pasos agigantados, cada vez más se puede notar que las relaciones (personales, comerciales, educativas, lúdicas, etc.) están migrando al mundo virtual. En base a esto, es conveniente conocer más profundamente algunas cifras que hagan conocer hasta qué punto los ciudadanos en el Ecuador hacen uso de la red.

El uso de la Internet aumenta vertiginosamente y día a día aumenta el número de usuarios en ella también se incrementa, en el caso de Ecuador estas aseveraciones se pueden confirmar y contrastar con los datos estadísticos otorgados por el Instituto Nacional de Estadísticas y Censos (INEC) y el ministerio de Telecomunicaciones y Sociedad de la Información, cabe destacar que no existe un consenso entre estas dos instituciones acerca del porcentaje real de uso de Internet en nuestro país, es por eso que se expondrán las cifras de las dos entidades.

El INEC indica que en un lapso de 10 años (2003-2013) en el Ecuador el uso de la Internet pasó de un mínimo 3% hasta un 40,4%, en ese mismo sentido, hasta finales del año 2013 el ministerio de Telecomunicaciones indica que un poco más del 50% de ecuatorianos tienen acceso a Internet en el país, ya sea con cuentas móviles o fijas.

La Agencia de regulación y control de las telecomunicaciones (ARCOTEL), afirma que hasta Marzo del 2014, en Ecuador el acceso al Internet se realiza en un 62,97% mediante conexiones fijas y el 37,03% restante mediante conexiones móviles (EduTic Ecuador 2014), pero las cifras no puntualizan si las cuentas son de personas particulares o de empresas públicas o privadas.

En lo que respecta al acceso a Internet en zonas urbanas, el INEC (2013) indica que hasta finales del 2013 el 47,6% de la población urbana tuvo acceso a la red, mientras que en las zonas rurales tan solo el 25,3% de dicha población accedió a Internet.

A continuación se muestra información acerca del acceso a Internet en las diferencias provincias del país, los datos de igual forma han sido tomados del INEC:

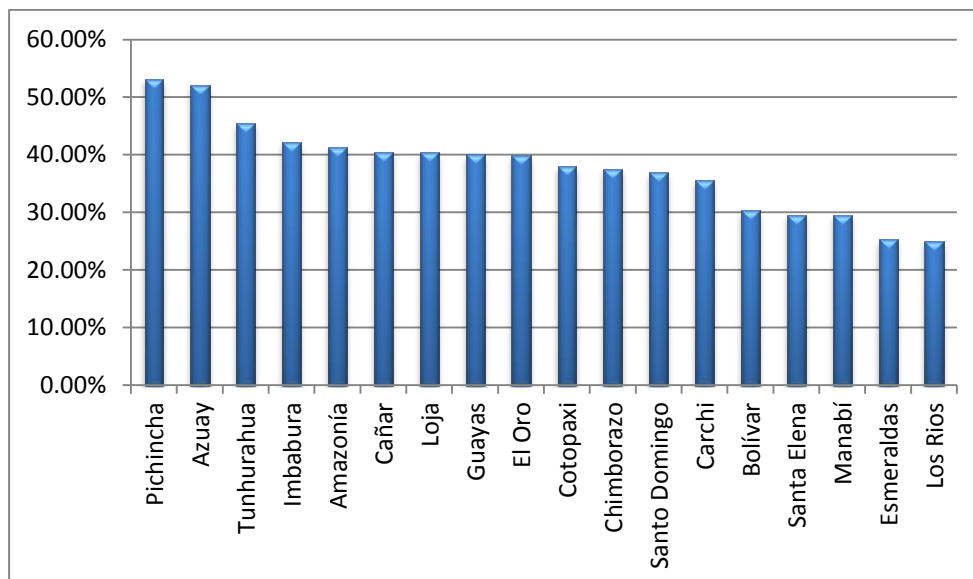


Figura 1. Acceso provincial a Internet en el Ecuador, hasta Diciembre 2014

Fuente: INEC

Elaboración propia

Hasta Diciembre del 2014 la provincia con mayor acceso a Internet fue Pichincha con un 53,10% de su población, la provincia de Loja se ubica séptima con un 40,40% mientras que Los Ríos se convierte en la provincia con menor acceso a Internet con un 25%, cabe destacar que el estudio se basa en el acceso a Internet durante los doce últimos meses a la aplicación de la encuesta (INEC, 2013).

2.2. Brecha digital: definiciones generales.

En este apartado se explicará a profundidad la definición de brecha digital y sus implicaciones en el uso de Internet. El uso de la red está sujeto, como cualquier otro servicio, a ciertas limitantes que hacen que no cualquier persona la pueda usar. Para hacer más técnica la explicación del significado de brecha digital se tomará el concepto otorgado por Yu (2002) quien afirma que:

La brecha digital comprende una amplia gama de tecnologías de comunicación y múltiples capas de acceso y uso de estas tecnologías, es un problema multifacético que abarca una gran variedad de desigualdades y diferencias basadas en: raza, género, edad, ingresos, educación, tipo de hogar, ubicación geográfica, habilidades físicas y el nivel de desarrollo económico del país en que se encuentre (p. 7).

En ese mismo sentido Álvarez (2011) intenta determinar el porqué de la existencia de dicha

brecha argumentando lo siguiente: “la brecha digital surge porque existen personas con acceso limitado o nulo a Internet, mientras que otras tienen un acceso pleno al mismo, lo que genera inequidad” (p. 57). Cabe destacar que el término latino *Brecha Digital* proviene del término anglosajón *Digital Divide*; el cual, algunos autores afirman que se usa desde principios de los años 90 (Aladi, 2003; Gunkel, 2003 y Vázquez, 2014).

Para ahondar más en este tema Cooper (2005) propone la existencia de categorías para determinar la conectividad y acceso a Internet de una persona, sus conclusiones se presentan a continuación:

1. *Plenamente conectado*: se refiere a aquellas personas que poseen planes de Internet fijos o móviles contratados con proveedores que ofrecen conexiones de alta velocidad y que gracias a dichos planes siempre están conectados.
2. *Parcialmente conectado*: son aquellas personas que poseen un servicio básico o limitado de Internet en sus casas o en sus móviles, el cual que no ofrece la velocidad y conectividad adecuada.
3. *Potencialmente conectado*: en este grupo están las personas que a pesar de poseer un computador o un celular, no tienen un proveedor que les ofrezca el servicio, por lo tanto deben buscar otro medio para conectarse.
4. *Desconectado*: son aquellas personas que no poseen conexión a Internet y tampoco computadora o un celular (p. 32).

Las categorías elaboradas por Cooper ayudan a evidenciar la existencia de la brecha digital, desde las personas que se conectan cuando desean hasta las personas que no poseen ningún medio para acceder a la red. Acerca de este tema Álvarez (2011) afirma que “la desigualdad que existe fuera de Internet poco a poco se traslada a ella, debido a que Internet no es una fórmula mágica para erradicar la pobreza” (p. 58).

Se puede considerar a la brecha digital como un problema social, que involucra aspectos socioeconómicos, culturales y sociales, donde el factor que define las desigualdades o diferencias es el acceso y uso de la tecnología.

Continuando con el análisis de las implicaciones de la brecha digital Floridi (2007) afirma que:

La brecha digital puede convertirse en un abismo y generar nuevas formas de discriminación, entre los que tienen acceso a Internet (ricos en información) y los que no poseen ese acceso (pobres en información). El uso de las TICs podría convertirse en otra forma de agravar las

desigualdades, toda vez que la revolución digital no reduce las barreras económicas, personales y sociales (p. 16).

Como se puede observar en todas las definiciones y aportes sobre el significado de brecha digital, apuntan a un mismo tema: la desigualdad; la cual se genera por el obvio motivo de que no todas las personas están en capacidad de acceder a la red cuando ellas lo deseen. A continuación se presentan datos y cifras del caso ecuatoriano.

El índice “**Networked Readiness**” publicado por el Foro Económico Mundial (2014), mide en una escala desde 1 (el peor) a 7 (el mejor), el rendimiento de 148 economías en el aprovechamiento de las tecnologías de información y comunicación para impulsar la competitividad y el bienestar, como se puede observar en la siguiente Figura el Ecuador ocupa el puesto 82 con un índice de 3,85.

Rank	Country/Economy	Value	2013 rank (out of 144)	Rank	Country/Economy	Value	2013 rank (out of 144)
1	Finland	6.04	1	75	Romania	3.95	75
2	Singapore	5.97	2	76	Sri Lanka	3.94	69
3	Sweden	5.93	3	77	Moldova	3.89	77
4	Netherlands	5.79	4	78	Philippines	3.89	86
5	Norway	5.70	5	79	Mexico	3.89	63
6	Switzerland	5.62	6	80	Serbia	3.88	87
7	United States	5.61	9	81	Ukraine	3.87	73
8	Hong Kong SAR	5.60	14	82	Ecuador	3.85	91
9	United Kingdom	5.54	7	83	India	3.85	68
10	Korea, Rep.	5.54	11	84	Vietnam	3.84	84

Figura 2. Índice Networked Readiness 2014.

Fuente: Foro Económico Mundial, recuperado el 16/06/2015.

Elaborado por: Foro Económico Mundial.

2.2.1. Aspectos de la brecha digital.

Como ya se ha venido hablando en temas anteriores la brecha digital se refiere a la desigual o dispar capacidad que tienen las personas, empresas, entidades países, regiones, etc., para acceder a las TICs. En los apartados que vienen a continuación se profundizará acerca de los aspectos que engloban a la llamada brecha digital en estos hoy por hoy, se hace relación al uso de las herramientas tecnológicas.

2.2.1.1. *Brecha de calidad de uso.*

Para lograr una mejor comprensión de este tema es conveniente valerse de la conceptualización otorgada por Camacho (2007) quien explica que:

La brecha de calidad de uso se basa en la limitación o falta de posibilidad que las personas tienen para poder acceder a las diferentes herramientas que proporciona la red. Un ejemplo clave para comprender la situación, en la que se basa la brecha de calidad de uso; sería el no tener conocimiento de la existencia o la forma de usar recursos en la red, como por ejemplo herramientas de medicina, comercio, información redes sociales etc., lo que implica una desventaja para quienes necesiten solucionar problemas (párrafo 4).

La brecha de calidad de uso está fundamentada en que, a pesar que la persona posea acceso a la red, ésta no está en la capacidad de utilizar o simplemente no conoce de la existencia de herramientas que ofrece Internet para problemas cotidianos o profesionales. Obviamente esta es una gran desventaja que ciertas personas tienen que afrontar y que a la postre se convertirá en una disminución de las competencias tecnológicas.

2.2.1.2. *Brecha de acceso o conectividad.*

Camacho (2007) explica que “la brecha de acceso se basa en la posibilidad o bien la dificultad que una persona puede llegar a tener para disponer de una computadora conectada al Internet, esto de igual forma incluye lo que son las redes avanzadas así como tener servidores funcionales” (párrafo 2). Para contrastar este concepto se presente la definición de Castells (2001) quien afirma que la brecha de accesibilidad es la diferencia entre quienes tienen Internet y los que no.

La UNESCO, en su estrategia a mediano plazo (2002- 2007), concluye que el acceso al conocimiento, a la comunicación y a la participación genuina en la sociedad del conocimiento está impedida para la mayor parte de la humanidad por una variedad de obstáculos que se detallan a continuación:

- Infraestructura inadecuada, tanto con la relación de disponibilidad energética y de hardware, como por las instalaciones de telecomunicaciones y de distribución masiva.
- Recursos financieros insuficientes.
- Analfabetismo persistente.

- Ausencia de un ambiente científico que propicie el uso de dichas tecnologías.
- Falta de entrenamiento en el uso de las TICs.
- Mantenimiento o asistencia no confiable.
- Contenidos que no son apropiados o incluso imposibles de usar por razones lingüísticas y culturales (p.174).

La brecha de acceso es simplemente la incapacidad que tienen: personas, entidades, países y regiones para poder acceder efectivamente a la red, ya sea por motivos: tecnológicos, económicos, educativos, sociales o culturales.

2.2.1.3. Brecha de uso.

Continuando con los tipos de brechas digitales existentes, en este apartado se conocerá en que consiste la llamada brecha de uso, para ello se citará a Camacho (2007), quien explica que:

La brecha de uso se basa en la incapacidad o dificultad que una persona tiene para usar las tecnologías existentes, ya que se ha encontrado que hay una gran diferencia entre llegar a tener acceso a la tecnología y saber usarla (párrafo 3).

Ahondando más en este tema Dimaggio, Hargittai, Neuman, y Robinson (2004), afirman que “la disparidad en el uso se fundamenta en que hay personas que saben usar las tecnologías y hay personas que no saben hacerlo” (p. 2).

Actualmente, tanto en nuestro país como en cualquier otro, existen entidades públicas y/o privadas que se dedican a estudiar y analizar estadísticas relacionados al tema de uso de tecnologías, y con los resultados obtenidos se ha llegado a acuñar el término analfabetismo digital, el cual se refiere a grado de desconocimiento del uso de tecnologías, Álvarez (2011). En el caso ecuatoriano y más específicamente en lo que se refiere a la brecha de uso, el INEC 2013 considera a una persona como *analfabeta digital* cuando cumple simultáneamente tres características: “1) no tiene celular activado, 2) en los últimos 12 meses no ha utilizado computadora, 3) en los últimos 12 meses no ha utilizado Internet” (p. 26).

2.2.2. Causas o factores que influyen en la brecha digital.

Tanto en el caso de brecha digital como en el de cualquier otro problema social, económico, cultural o intelectual, existen causas o factores que motivan su existencia, en este apartado se recogerán las principales causales que han desencadenado la brecha digital y así mismo se

explicará cada una de ellas.

Torres (2012) realiza un favorable aporte en lo que respecta a las determinantes de la desigualdad digital, a continuación su aporte:

La búsqueda de relaciones que tienen el nivel de ingresos como variable determinante, no se centra solamente en la posibilidad de contar con un equipo y una conexión, su incidencia va más allá y actúa conjuntamente con otras variables como el nivel de educación, el contexto familiar y el contexto social; no es posible concebir las relaciones de forma aislada sino de manera sistémica, puesto que, una sola variable no describe apropiadamente la desigualdad digital (p. 59).

A continuación se presentan las principales determinantes de la desigualdad digital.

2.2.2.1. Factor educativo.

Fácilmente se puede interpretar que una de las principales causas del no acceso a las tecnologías viene dado por deficiencias a nivel educativo, en base a esto Torres (2012) indica que:

El nivel de educación determina muchas de las diferencias entre los usuarios de tecnología. Esto se debe en gran medida a que una persona con educación formal, transita por un conjunto de experiencias y procesos que toman un tiempo considerable, en el que a más de asimilar contenidos se desarrollan relaciones que hacen a la persona partícipe de una cultura. Como consecuencia el usuario representa un perfil de necesidades, habilidades y posibilidades que determinan un comportamiento frente a la tecnología (p. 60).

Para entender mejor la relación educativa con la brecha digital, Gutiérrez y Gamboa (2010) realizaron un estudio en Colombia, México y Perú, gracias a ello determinaron que “en personas de bajos niveles de ingreso, la educación es el principal determinante del nivel de digitalización y ante la falta de posibilidades de acceso es la necesidad educativa la que fomenta el desarrollo de habilidades y destrezas tecnológicas” (p. 345). De esto se interpreta que, en los países en los que se llevó a cabo el estudio, el acceso a las TICs está ligado al nivel de educación y la educación a su vez, al nivel de ingresos.

Al hablar de cómo el factor educativo influye en la brecha digital, es conveniente citar el estudio de Howard, Rainie, y Jones (2001), en el que buscaron determinar las actividades que realizan

en Internet ciertos usuarios dependiendo de su nivel de educación, a continuación sus resultados:

- En base al nivel de escolaridad se encontraron diferencias significativas determinadas por el nivel de educación.
- Los usuarios con título universitario presentaban mayor predisposición a la búsqueda de información sobre salud, finanzas, fuentes de trabajo y noticias en general, en mayor grado que aquellos usuarios con niveles de formación educativos más bajos.
- Las personas con nivel educativo más bajo presentaron mayor predisposición a las actividades de ocio y diversión en línea.

Graham, (2008) e Iske, Klein, y Kutscher (2005), sentencian que el factor que presenta mayor impacto en el uso del Internet es el nivel educativo tanto para actividades de comunicación como de búsqueda de información. A mayor grado de escolaridad de la persona el uso de la Internet desemboca en actividades más avanzadas y de índole profesional, mientras que a menor grado de escolaridad el internauta se enfoca a actividades meramente recreativas.

El Pew Research Center, es una organización sin fines de lucro que realiza estudios acerca del uso y conocimiento de las TICs en los Estados Unidos y otros países anglosajones. El Pew Research Center hasta el año 2014¹, determinó que el acceso a Internet o a correo electrónico por medio de dispositivos móviles en los estudiantes de ese país está en las siguientes proporciones:

- 97% de los alumnos de posgrado acceden a Internet por este medio.
- 91% de los alumnos de pregrado lo hacen de la misma forma.
- 76% de los alumnos que cursan la secundaria o estudios inferiores lo hacen por medio de estos dispositivos.

En EEUU el acceso a Internet, más específicamente al de conexión móvil; se encuentra altamente masificado en sus estudiantes y de igual forma se puede apreciar que a mayor nivel educativo mayor es el acceso a la red.

Para poder tener una mirada más cercana al caso ecuatoriano, el Ministerio de Telecomunicaciones y de la Sociedad de la Información (MINTEL), indica que en el Ecuador desde el año 2012 hasta finales del 2014, el analfabetismo digital disminuyó del 21,4% al

¹ Información obtenida de: <http://www.pewInternet.org/data-trend/Internet-use/latest-stats/> recuperada el 17/06/2015.

14,4%², de igual forma el ministerio informa que actualmente el estado se encuentra implementando el Plan Nacional de Conectividad Escolar, con el cual se busca dotar al 100% de los establecimientos educativos de nivel primario y medio en todas las áreas urbanas y rurales del país con laboratorios de computación y acceso a Internet.

El factor educativo permite no solo determinar las personas que tienen o no acceso a la información; sino también, hace énfasis en saber qué hacer con los recursos y herramientas que ofrece el Internet para las actividades educativas; por tanto, se llega a la conclusión que para acortar la brecha digital no solo se debe tener en cuenta el aspecto tecnológico; sino también la información que se maneje y la adquisición de nuevos conocimientos.

2.2.2.2. Factores de ingresos.

Otro de los motivos que desencadenan la existencia de la brecha digital a nivel mundial es el factor de ingresos del internauta, para entender mejor este aspecto Hargittai (2010, citado por Torres, 2012) determina que:

Los ingresos constituyen el elemento diferenciador para contar con equipamiento de altas prestaciones, varios puntos de acceso y conexiones de banda ancha. Un equipo de gran capacidad permite ejecutar aplicaciones que requieren mayor procesamiento, lo que puede incidir en los tipos de herramientas que se utilicen; un determinante del tiempo de conexión es la cantidad de puntos de acceso de los que dispone el internauta, ya sean estos móviles para teléfono u ordenador portátil o una conexión dedicada desde el hogar; la posibilidad de conectarse desde casa y/o a través de una conexión móvil a más de la otorgada por la institución educativa o por el lugar de trabajo, incrementa el tiempo en línea y las habilidades y destrezas. (p. 61)

Resulta muy acertada la conclusión de Hargittai, debido a que, como se ha analizado con anterioridad; no basta solamente con que la persona tenga los conocimientos referentes al uso de Internet, sino que también requiere un equipo de cómputo o un dispositivo móvil con acceso permanente a la red (de preferencia con conexión del alta velocidad), para poder decir que dicha persona no se encuentra excluida del mundo digital y como es obvio dichos recursos demandan tener un cierto nivel de ingresos que no todas las personas lo tienen.

A nivel global, la Unión Internacional de Telecomunicaciones (UIT)³ ha determinado que, hasta el año 2012, el **precio** del acceso a Internet se convierte en la mayor limitante que encuentran

² Información obtenida de <http://www.telecomunicaciones.gob.ec/conectividad-escolar/> recuperada el 17/06/2015

³ En base al estudio: "Medición de la sociedad de la información 2012".

los usuarios al momento de acceder a la red. De igual forma la UIT ha determinado que a escala mundial los servicios de telecomunicaciones y de Internet están empezando a abaratare, esto debido a que, según sus propios cálculos, en relación a la canasta básica mundial (la cual incluye el costo medio de los servicios de telefonía fija, Internet de banda ancha fija y móvil celular); entre los años 2008 y 2011 el precio del acceso a las TICs disminuyó un 30% mundialmente y la reducción más dramática fue la de los servicios Internet de banda ancha fija, cuyos precios bajaron en promedio un 75%.

En el estudio sobre “Los factores socioeconómicos que afectan a los patrones de uso de Internet doméstico en la central Queensland” se encontraron coincidencias entre los niveles de ingresos y los tiempos de uso de Internet, dando como resultado que quienes pertenecen a los niveles de ingreso mayores permanecen más tiempo conectados y por lo tanto pueden desarrollar más conocimientos y experiencia, Luna (2012, cita a Taylor, Zhu, Dekkers, y Marshall, 2003) (p. 6).

Para el caso latino americano Hilbert (2010), afirma que los factores que inciden en el acceso a Internet son similares a los de los países desarrollados, independientemente de la presencia de los diferentes niveles de inequidad que existe en la distribución de los ingresos. Es decir, que los países en vías de desarrollo poseen las mismas restricciones de acceso a la red que los países desarrollados.

Para el caso ecuatoriano se han encontrado datos importantes en la tesis de grado realizada por Muñoz (2013) quien determina, en base a los ingresos de las familias en el año 2010, que “el quintil⁴ 5 de la población del Ecuador presenta la mayor tasa de uso de Internet con 53,8% frente al 13,2% de uso de las personas ubicadas el quintil 1” (p. 63). Lo que quiere decir que en el Ecuador las personas que más acceden a Internet son quienes poseen mayores ingresos, lo que resulta en la exclusión de las personas ubicadas en los quintiles más bajos de la economía.

En la siguiente gráfica se evidencia la evolución del acceso a Internet en el Ecuador por parte de los quintiles de su población, durante los años 2008, 2009 y 2010.

⁴ Quintil es un término estadístico que se usa para referirse a las características que se encuentran en un quinto de la población estudiada, es decir, si se habla del quintil 1 se refiere a las características de la primera quinta parte de la población estudiada y así sucesivamente con los restantes cuatro quintiles.

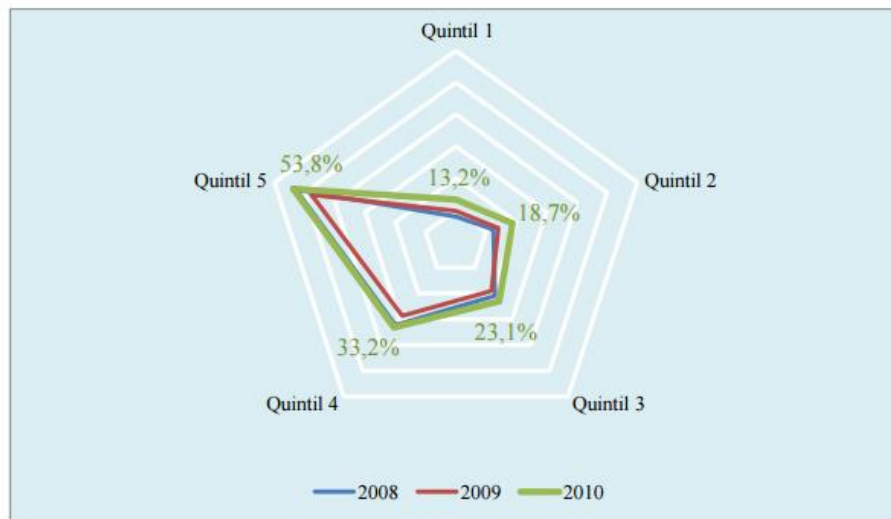


Figura 3. Uso del Internet en el Ecuador por quintiles de ingreso 2008-2010
Fuente: Muñoz (2010, p. 64)
Elaborado por: Muñoz (2010, p. 64)

Durante el período 2008 - 2010 en el Ecuador, en base a los quintiles de ingresos, quienes tuvieron mayor acceso a Internet son las familias con ingresos más altos y dicho acceso se reduce conforme se reducen los ingresos de las familias, por lo tanto y en base a este estudio se puede colegir que el acceso a la red de los habitantes del Ecuador no es directamente proporcional a su nivel de ingresos

2.2.2.3. Factor género.

A priori se podría pensar que el género de la persona interesada en el acceso a las TICs no debería ser un factor determinante en el acceso, para argumentar esta aseveración el Pew Research Center en el año 2010 determinó que en EEUU el 78% de su población masculina tuvo acceso a Internet, mientras que el 76% de su población femenina también tuvo acceso, lo que hace notar que el margen es mínimo y su impacto no es significativo.

En el caso ecuatoriano no existen estudios que ayuden identificar la brecha digital por el género de la persona; sin embargo, gracias al último censo realizado en el año 2010, el INEC determinó los siguientes porcentajes de analfabetismo digital por género:

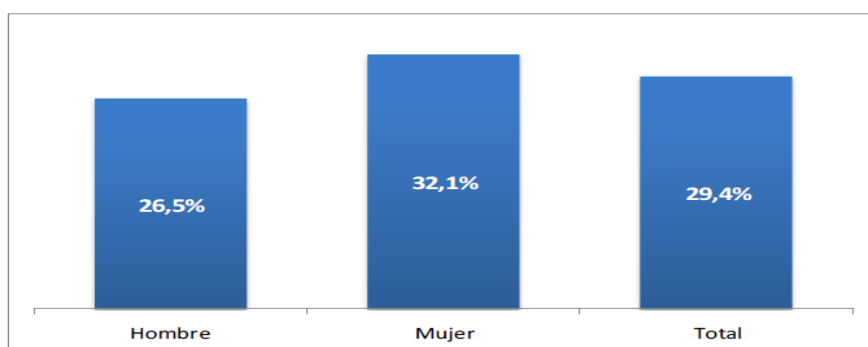


Figura 4. Analfabetismo digital por género en el Ecuador, según el censo 2010.

Fuente: Muñoz (2010, p. 74)

Elaborado por: Muñoz (2010, p. 74)

2.2.2.4. Factor étnico

Calcagno y Hernández (2003) aseveran que “la pertenencia a una etnia es otro factor de la brecha digital latinoamericana y caribeña.” Según el estudio realizado por Chakraborty y Bosman (2005) realizaron la comparación sobre la tenencia de un computador entre Afro-Americanos y Europeo-Americanos, los resultados indicaron que “el 40% de los hogares de familias caucásicas (blancas) con menores ingresos tienen computador, a diferencia del 27% de las familias Afro-Americanas más pobres” (p. 45), lo que quiere decir que incluso comparando familias de bajos ingresos pero de distinta etnia se encuentran diferencias significativas en aspectos digitales. Para el caso de Ecuador, en lo que respecta a la brecha digital por etnia gracias al censo de población y vivienda del año 2010, el INEC determinó lo siguiente en este aspecto:

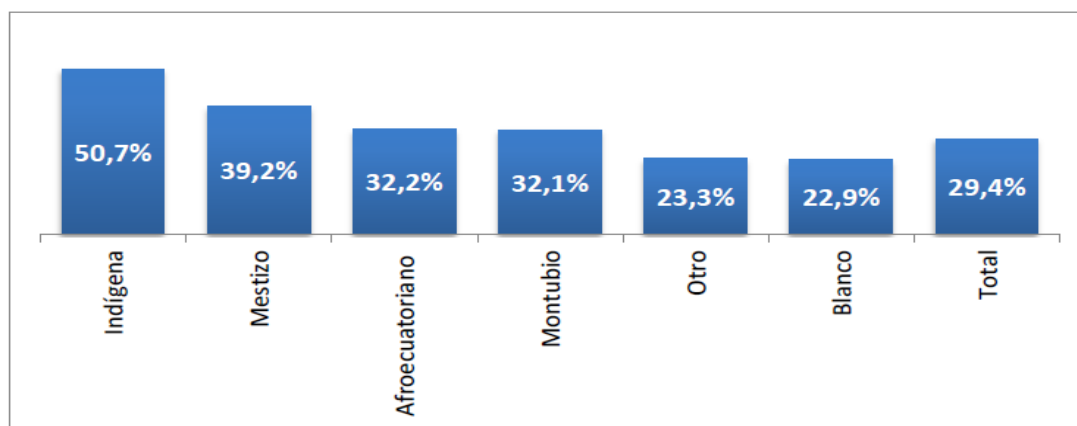


Figura 5. Analfabetismo digital por auto-identificación étnica en el Ecuador, hasta el año 2010.

Fuente: Muñoz (2010, p. 79)

Elaborado por: Muñoz (2010, p. 79)

2.2.2.5. *Factor geográfico.*

Serrano y Martínez (2003) afirman que “las grandes ciudades tienen más acceso a las tecnologías de la información que las ciudades pequeñas y áreas rurales” (p.75). La Comisión Europea (2006, citado por Álvarez, 2011) afirma:

La brecha entre zonas rurales y urbanas no distingue entre países desarrollados y en vías de desarrollo, pese al aumento general de la conectividad de banda ancha, el acceso sigue siendo limitado en las regiones más apartadas y rurales a causa de los altos costos que implican llevar conectividad a una escasa y lejana población (p. 71,72).

Para el caso de Ecuador el INEC, en base al censo del año 2010, determinó que el 40,2% de las personas que viven en zonas rurales son consideradas analfabetas digitales, un alto porcentaje comparado con el 23,2% de las personas que viven en zonas urbanas.

Para paliar esta situación el MINTEL desde el año 2010 puso en marcha el llamado “Plan Infocentros⁵”, en dicho plan se proyecta la creación de 1200 de dichos establecimientos a nivel nacional, los cuales se ubicarán en las zonas marginas de los sectores urbanos y rurales. El portal web www.infocentros.gob.ec⁶ reporta que hasta Mayo del 2015 el Ecuador cuenta con 490 infocentros y un megainfocentro en Tumbaco. En el mismo portal se informa que gracias a estos centros se ha logrado beneficiar a aproximadamente 2´886.591 ciudadanos, con una inversión que ronda los 19 millones y medio de dólares, de igual forma se ha logrado capacitar en TICs a aproximadamente 185.406 personas. Es evidente que con la implementación de estas políticas públicas el estado busca reducir la brecha digital en base a la locación geográfica de las personas y a su nivel de ingresos.

2.2.3. *Desigualdad digital en América Latina y el Ecuador.*

El organismo internacional más importante en temas de acceso a las TICs es la (UIT), quien publica regularmente estadísticas de este tema a nivel mundial, en su informe del año 2012 determinó los siguientes datos para la región de América Latina y el Caribe (se han recogido los datos más relevantes del estudio):

⁵ Los Infocentros son espacios comunitarios de participación y desarrollo, que garantizan el acceso inclusivo a las Tecnologías de la Información y Comunicación de las poblaciones de parroquias rurales y urbanas marginales del Ecuador (MINTEL).

⁶ Consultado el 17/06/2015.

- A finales del año 2011, 20 (de 33) países de la Región de América Latina y el Caribe registraban más abonados a sistemas móviles-celulares que habitantes, entre los que figuraban Argentina, Brasil, Chile, Ecuador, Guatemala, Panamá, Perú y Uruguay.
- En 2011, se sumaron más de 30 millones de abonados a sistemas móviles-celulares en la Región de América Latina y el Caribe, la mayoría de ellos en Brasil (20 millones) y México (3 millones).
- El porcentaje de usuarios de Internet sigue creciendo en América Latina y el Caribe, y en 2011 llegó al 39%.
- En América Latina y el Caribe el porcentaje de hogares con acceso a Internet sigue aumentando, llegándose al 29% de hogares en 2011.
- El porcentaje de hogares con computadora está aumentando en América Latina y, a finales de 2011, casi el 36% de los hogares de la Región de América Latina y el Caribe contaba con una computadora.

Tabla 1. Posiciones de países latino-americanos dentro del Networked Readness Index 2014

PAÍS	POSICIÓN 2014	POSICIÓN 2013	PAÍS	POSICIÓN 2014	POSICIÓN 2013
Panamá	43	46	El Salvador	98	93
Costa Rica	53	53	Argentina	100	99
Uruguay	56	52	Guatemala	101	102
Colombia	63	66	Paraguay	102	104
Brasil	69	60	Venezuela	106	108
México	79	63	Honduras	116	109
Ecuador	82	91	Bolivia	120	119
Perú	90	103	Nicaragua	124	125
República Dominicana	93	90			

Fuente: Foro Económico Mundial
Elaboración propia

El país latino-americano mejor posicionado dentro del ranking 2014 es Panamá, quien ocupa la casilla 43 dentro de un total de 148 economías a nivel mundial y en la última casilla la ocupa Nicaragua quien se ubica en la posición 124. De igual forma se puede evidenciar que los saltos más cualitativos los han dado Perú y Ecuador, quienes han mejorado 13 y 9 posiciones respectivamente. En contraste se encuentran México y Brasil quienes cayeron 16 y 9 posiciones respectivamente.

Estos valores indican que como región aún falta un largo camino por recorrer para poder erradicar la brecha digital que nos separa de los países desarrollados.

En el caso de Ecuador el MINTEL y la ARCOTEL reportan trimestralmente el acceso a las TICs de los habitantes del país, hasta marzo del 2015, se pueden evidenciar las siguientes estadísticas:

- En el Ecuador existen 6´549.733 abonados a un proveedor de Internet lo que representa el 40,71% del total de la población.
- En el Ecuador se registran 14´043.191 personas que son usuarias⁷ de Internet, lo que representa el 87,28% de los habitantes.

Otra de las estrategias gubernamentales empleadas para paliar la brecha digital existente en nuestro país, es el llamado “*PLAN NACIONAL DE BANDA ANCHA*”, el MINTEL afirma que dicho plan busca “generar condiciones de uso del espectro radioeléctrico para adaptar nuevas formas en la industria que proporciona servicios inalámbricos, garantizando el uso eficiente del espectro y la sana competencia, para satisfacer la demanda de los servicios y promover la reducción de precios” (párrafo, 1). Los objetivos que se busca obtener son:

- Mejorar la calidad de vida de los ecuatorianos mediante el uso, introducción y apropiación de las nuevas tecnologías de información y comunicación.
- Decremento los precios de acceso al servicio de Internet de banda ancha
- Impulsar el despliegue de redes y servicios a nivel nacional.
- Permitir a todos los ecuatorianos independientemente de su condición socio-económica y ubicación geográfica el acceso a los servicios de banda ancha de calidad (párrafo, 2).

En base a las cifras expuestas, se evidencia que en América Latina la brecha digital tiende a disminuir. Ecuador muestra un evidente escalamiento de posiciones en el ranking de disminución de la desigualdad digital, mejorando año a año su posición; sin embargo, se considera oportuno seguir trabajando constantemente con la finalidad de ir mejorando la accesibilidad a las TICs, consiguiendo el impacto en las mismas; como también, incrementando la infraestructura y el uso de las tecnologías de manera que la brecha digital desaparezca poco a poco.

2.3. El rendimiento académico y el uso del Internet.

En este apartado se buscará determinar la incidencia del uso de Internet en el rendimiento académico de quien lo use como herramienta de estudio, para ello primeramente se debe

⁷ El MINTEL y ARCOTEL afirman que usuario es la persona que hecho uso del Internet al menos una vez, sin importar si dicha persona tiene o no una cuenta de Internet, posee o no celular y/o computadora; es decir no se determina como accedió a la red.

conocer que es y que implica el rendimiento académico, para lo cual se tomará la definición otorgada por Lucas (1998, citado por Torres, 2012):

El rendimiento académico constituye el nivel de éxito alcanzado por parte del estudiante en sus actividades académicas. Usualmente se lo relaciona de forma directa con la inteligencia, pero en realidad depende de una serie de factores como la familia, el ambiente escolar e inclusive algunos hábitos asimilados a tempranas edades como el interés por la lectura; también pueden incidir factores como el nivel socioeconómico, del que dependen los medios de los que el estudiante dispondrá y algunas capacidades elementales como la comprensión y producción lingüística (p. 72).

Por otro lado, Tamez (2012) indica que “el rendimiento académico es una medida de las capacidades del alumno, que expresa lo que éste ha aprendido a lo largo del proceso formativo” (p. 32).

El rendimiento académico es la medida del éxito que puede alcanzar un estudiante en sus actividades académicas, y que no es el simple resultado de la inteligencia de una persona, sino que también intervienen diversos factores personales, sociales y económicos, e incluso el educativo durante toda su formación académica.

López (2013) indica que “con la web 2.0 el acto educativo se transforma; se transita de un medio pasivo a uno activo en el que los alumnos se convierten en autores de nuevas ideas más que en consumidores de datos e información” (p. 105). Como se puede observar en este primer aporte, Internet se ha vuelto una herramienta para empoderar al estudiante y que deje de ser un ente pasivo en el proceso de aprendizaje.

La Fundación Chandra (2007) afirma que:

Con la presencia de herramientas tecnológicas, fáciles de usar, gratuitas y prácticas se pretende fortalecer la comunicación entre profesores y alumnos, gestionar el conocimiento y promover el desarrollo de las personas. Bajo este nuevo esquema de formación se apuesta por la innovación y la conformación de una nueva vida comunitaria así como el intercambio de ideas (p. 12).

Este aporte indica que las TICs fortalecen la comunicación entre profesores y alumnos y ayudan al desarrollo de las personas. En el mundo actual donde la innovación y las nuevas tecnologías aparecen a una velocidad vertiginosa, se debe tomar en cuenta que a más de las variables tradicionales, que han sido consideradas fijas en el rendimiento académico durante mucho

tiempo; ahora se deben sumar variables tecnológicas, que como ya se ha visto en los conceptos anteriores son vitales para el proceso de enseñanza-aprendizaje. Duart, Gil, Puyol, y Castaño (2008) define a las tecnologías actuales como un factor importante del rendimiento académico.

Acerca de la incidencia directa del uso de Internet sobre el rendimiento estudiantil Torres (2012,) indica que “el uso de Internet pueden tener un efecto indeterminado sobre el rendimiento de los estudiantes, en algunos casos incide positivamente, mientras que en otros, causa el efecto contrario” (p.74), para ello cita varias fuentes con estudios al respecto de este tema, los cuales han usado distintas variables para buscar relaciones entre el uso de Internet y el rendimiento académico, algunas de las variables usadas han sido:

- Las actividades que el usuario realiza en línea.
- La intensidad o tiempo de conexión que tiene el usuario.
- El contar con un computador y conexión a Internet desde el hogar.

Sin embargo, Torres (2012) señala que no se ha llegado a conclusiones determinantes, puesto que:

En distintos estudios con condiciones similares, los resultados han sido contradictorios; así mismo, en otros casos el uso de tecnología presenta efectos positivos en ciertas áreas cognitivas del usuario como en el desarrollo de habilidades espaciales y de memoria, la mejora de las habilidades de lectura, escritura y de procesamiento de información, pero estas no necesariamente se traducen en mejoras en el rendimiento académico (p. 75).

No necesariamente el uso de Internet incide positivamente en el rendimiento académico de una persona, se debe recordar lo expuesto en el concepto Lucas (1998), quien afirmaba que el rendimiento académico no es el resultado de una sola variable sino de varias que al ser conjugadas darán el resultado esperado.

Luna (2012) concluye que el uso de Internet no incide en el rendimiento académico de los estudiantes de carreras presenciales de la Universidad Técnica Particular de Loja, y sugiere “replantear el uso de tecnología como parte de las actividades de aprendizaje buscando un efecto positivo especialmente en la calificación que logra el estudiante” (p. 85).

En otros casos, como menciona Torres (2012) “el uso de tecnología puede presentar incidencia

positiva en ciertas áreas y negativa en otras, esto debido a que las ciencias en general y las asignaturas en particular tienen distinta naturaleza y forma de ser abordadas” (p. 77). Para reafirmar este punto cita el estudio de Antonijevic (2007) quien llegó a determinar que “el uso de computadores contribuye a un alto desempeño de los estudiantes en el área de ciencias y causa un efecto contrario en matemáticas” (p. 77). Estos resultados indican que la incidencia del uso de Internet en el rendimiento académico *depende*, de las variables, del acceso, del tiempo en línea, de lo que se esté buscando y como se lo busca, de los hábitos de lectura, etc.

2.3.1. El uso de las redes sociales y su relación con el aprendizaje.

En este apartado final se centrará en el análisis investigativo de la influencia que pudieran llegar a tener las redes sociales en el aprendizaje de una persona; por lo tanto, se iniciará otorgando la definición de red social. En su libro “Aprendiendo con Redes Sociales” Castañeda (2010) afirma que:

El término Red Social se refiere a aquellas herramientas telemáticas de comunicación que tienen como base la web, se organizan alrededor de perfiles personales o profesionales de los usuarios y tienen como objetivo conectar secuencialmente a los propietarios de dichos perfiles a través de categorías, grupos, etiquetados personales, etc., ligados a su propia persona o perfil profesional (p. 18).

Una red social viene a ser un medio de comunicación e interrelación de personas por medio de un sistema que usa Internet como base de su funcionamiento. De igual forma en el mismo libro se hace una aclaración muy importante, la cual afirma que, al momento de hablar de redes sociales éstas no se deben confundir con otros tipos de herramientas de comunicación en Internet. La autora afirma que a pesar de parecer redes sociales, no todas las herramientas comunicacionales lo son; es por ello que hace la siguiente división y también se presentan algunos ejemplos:

Medios de comunicación social (social media): conocidos como los medios más antiguos de la web 2.0, se usan para que los usuarios publiquen y compartan un determinado elemento de comunicación concreto y que se estructuran en torno a esos elementos. Ejemplos:

- Flickr.com
- Youtube.com
- Picassa.com

- Slideshare.com
- Vimeo.com
- Scrib.com

Medios de seguimiento a la actividad en la red (lifestreaming): son sitios donde el usuario se da de alta y publica información acerca de sus actividades en redes sociales con la finalidad de que otros usuarios puedan observar un compendio, en tiempo real, de todo lo que publica en la red el primer usuario. Ejemplos:

- Twitter.com
- Profilactic.com
- Identoo.com
- Jaiku.com
- Cliqset.com
- Friendfeed.com

Redes sociales (social networking sites): se refieren a herramientas más complejas donde la interacción entre usuarios es lo más importante y como se mencionó en el primer concepto tienen como objeto conectar sucesivamente a los propietarios de las cuentas a través de categorías, grupos etiquetados personales. Ejemplos:

- Facebook.com
- LinkedIn.com
- Hi5.com
- Myspace.com
- Tuenti.com
- Instagram.com
- Google+.com

Una vez que ya se ha identificado eficazmente el concepto y la clasificación de las redes sociales el siguiente paso es ahondar en la relación entre dichas redes y el aprendizaje, para ellos existen varias fuentes bibliográficas que ayudaran a entender el tema. Para iniciar se tomará el aporte de Gómez, Roses y Farías (2012), quienes afirman que:

Las redes favorecen la publicación de información, el aprendizaje autónomo, el trabajo en equipo, la comunicación, la realimentación, el acceso a otras redes afines y el contacto con otros expertos, entre otros elementos. Todo ello, tanto entre estudiantes en general, como entre el binomio estudiante y profesor; lo cual facilita el aprendizaje constructivista y el aprendizaje colaborativo (p. 132).

Esta afirmación indica que las redes sociales poseen un alto potencial en el ámbito educativo y además pueden promover el aprendizaje autónomo del estudiante y su relación con el docente.

Una conclusión valedera en este sentido, es la concedida por Valenzuela (2013) quien afirma que:

Las redes sociales ofrecen una gran gama de posibilidades para los docentes en términos de establecer interacción en diversas vías, empezando por colegas, estudiantes e incluso con las autoridades de sus instituciones. Sin embargo, siempre es conveniente definir inicialmente, qué es lo que se desea resolver con cierta herramienta tecnológica, pues con base en ello será más sencillo elegir qué solución pudiera ser la más adecuada; esto tomando en cuenta todos los factores involucrados en la situación de enseñanza a resolver (p.12)

Las redes sociales sirven de apoyo tanto a profesores como a alumnos, pero se debe tener presente el objetivo al que se quiere llegar con la red social para que no existan distorsiones, ni malos usos.

Ya se ha hablado de que las redes sociales pueden favorecer tanto a los profesores como a los alumnos en el proceso de enseñanza-aprendizaje pero valdría tomar en cuenta las palabras de Aparici (2010, citado por Pérez García, 2013) quien afirma que:

El alumnado actual pertenece al grupo de nativos digitales que han nacido inmersos en la vorágine tecnológica y que dominan desde muy pequeños cualquier recurso digital, por lo que la facilidad de adaptación a cualquier innovación tecnológica que se introduzca en su día a día escolar es muy superior a la que había hace un par de décadas (p.12).

Esta es otra conclusión que alienta el uso de las redes sociales como parte de un nuevo modelo de enseñanza. Un contrapunto a tomar en cuenta en este aspecto, vendría a ser que no todas las redes sociales o no todas las herramientas web existentes sirven para ayudar al desempeño educativo, para entender mejor esta idea, De Haro (2010) concluye:

Las redes sociales más interesantes para la educación son aquellas que permiten crear redes a los usuarios independientes del resto de miembros del servicio, de forma que los profesores y alumnos puedan disponer de un espacio único para ellos sin injerencias exteriores de ningún tipo. las redes que cumplen dichas características son: a) SocialGO, b) Grouply, c) Ning, d) Gruo.ps, e) Tweducate, f) Edmodo y g) Shoutem (p. 7).

En el mismo trabajo De Haro hace referencia al término *redes sociales estrictas* y explica que “este tipo de redes son las que presentan un mayor valor en su aplicación educativa debido a su inespecificidad, con lo que se pueden adaptar libremente según las necesidades” (p. 9). Es decir que las redes sociales estrictas son aquellas que parecieran que no poseen un fin en específico y que por lo tanto el usuario o los administradores puedan darle el uso que más les convenga.

De Haro recomienda, para el ámbito educativo, que se distinga entre las redes sociales que se pueden descargar de Internet y que se pueden ejecutar en los servidores propios del centro educativo versus aquellas que están alojadas en servidores de terceros en manos de empresas especializadas; el motivo viene dado debido a que las primeras poseen la ventaja de tener los datos en los ordenadores de la propia institución educativa y de tener un control total sobre los datos que tienen relación con la red social.

A continuación se presentan las ventajas específicas que poseen las redes sociales en el ámbito educativo, las cuales han sido desarrolladas por De Haro (2010):

- Permite centralizar en un único sitio todas las actividades docentes, profesores y alumnos de un centro educativo.
- Aumento del sentimiento de comunidad educativa para alumnos y profesores debido al efecto de cercanía que producen las redes sociales.
- Mejora del ambiente de trabajo al permitir al alumno crear sus propios objetos de interés, así como los propios del trabajo que requiere la educación.
- Aumento en la fluidez y sencillez de la comunicación entre profesores y alumnos.
- Incremento de la eficacia del uso práctico de las TIC, al actuar la red como un medio de aglutinación de personas, recursos y actividades. Sobre todo cuando se utilizan las TIC de forma generalizada y masiva en el centro educativo.
- Facilita la coordinación y trabajo de diversos grupos de aprendizaje (clase, asignatura, grupo de alumnos de una asignatura, etc.) mediante la creación de los grupos apropiados.
- Aprendizaje del comportamiento social básico por parte de los alumnos: qué puedo decir, qué puedo hacer, hasta dónde puedo llegar, etc. (p. 22).

Para finalizar, Naso, Balbi, Di Gracia y Peri, (2012) exponen algunos ejemplos con los cuales los docentes, usando las redes sociales como herramientas y al mismo tiempo como el medio natural para llegar a sus alumnos, pueden promover el aprendizaje en sus alumnos:

- Usar o crear una aplicación propia en Facebook o una wiki para realizar trabajos grupales.
- Invitar a estudiantes que asistan a una clase via streaming.
- Responder preguntas, dejar tareas y anunciar fechas de eventos y exámenes en Facebook y/o Twitter.
- Compartir libros, pdfs, videos y otro material interesante usando una wiki o creando su comunidad virtual.
- Seguir a profesores, compañeros de clase y otros expertos en un tema via twitter, slideshare o RSS.
- Tener apuntes de la asignatura en un blog, wikis (p. 12).

El uso de las redes sociales puede ser considerado como una herramienta de apoyo alternativo al proceso de enseñanza-aprendizaje, aprovechando la ventaja de ser ampliamente utilizada por los estudiantes y los docentes, ayudándoles a tener mejor interacción entre compañeros de clase e incrementando la confianza con el docente. Es necesario enfatizar que toda red social debe ser manejada con precaución para evitar situaciones incómodas que vayan en contra de la integridad del usuario.

2.3.2. El uso de las redes sociales y su incidencia en el rendimiento académico

En el apartado anterior se habló acerca de la relación existente entre el uso de redes sociales y la enseñanza, a hora se profundizará en como dicho uso de redes sociales incide en el rendimiento académico, para ello se citarán estudios muy específicos acerca del tema. La información se presentará otorgando las conclusiones de cada estudio.

En su estudio a los alumnos de nivel medio superior de la ciudad de Chihuahua-México, González, Soltero y González (2012) concluye los siguientes puntos:

1. El bajo rendimiento académico, debido a la adicción en la red social de Facebook, afecta a por lo menos dos de cada cinco estudiantes de la Preparatoria 20 de la UANL.
2. Se confirma que las redes sociales contribuyen en el aprovechamiento académico al acceder a los contenidos de profesores e investigadores e interactuar con ellos (p. 20).

Existen puntos positivos y negativos en el uso de redes sociales, y lo más importante es la

conclusión de que al menos 2 de cada 5 alumnos poseen bajo rendimiento académico debido al uso excesivo de redes sociales.

Otro estudio relevante en este tema es el realizado por Santos (2010), dicho estudio investiga la importancia conjunta de la estructura y el contenido de las redes sociales de amistad de estudiantes de secundaria para su rendimiento académico, se debe tomar en cuenta que se usaron datos del Estudio Nacional Longitudinal de Salud Adolescente (National Longitudinal Study of Adolescent Health) de los Estados Unidos. El autor concluye:

1. Se encontró que estudiantes situados en redes sociales altamente cohesionadas superan académicamente en matemáticas e inglés a estudiantes ubicados en redes poco cohesionadas, cuando ambos tipos de actores están conectados a estudiantes muy identificados con su escuela. Sin embargo; estudiantes situados en redes poco cohesionadas superan académicamente en las mencionadas asignaturas a estudiantes posicionados en redes altamente cohesionadas, cuando ambos tipos de actores están vinculados con estudiantes poco identificados con su escuela.
2. Los resultados sugieren que es crucial tener en cuenta tanto la estructura como el contenido de las redes sociales de los estudiantes para entender su rendimiento educativo (p. 7).

En dicho estudio se trató de buscar una relación entre el uso de redes sociales, amistad y rendimiento académico, como se ve en la primera parte de la conclusión 1, estudiantes que interactúan en redes sociales con alta interrelación de sus participantes superan académicamente en las materias de matemáticas e inglés a los estudiantes que interactúan en redes sociales con baja interacción de sus participantes pero esto sucede cuando los dos tipos de estudiantes se relacionan o tienen amistad con otros alumnos que se sienten muy identificados con su colegio.

Un estudio relacionado al tema se realizó en Chile a 5000 alumnos de nivel medio, el autor Halpern (2014) llega a las siguientes conclusiones relacionadas al uso de las redes sociales y el rendimiento académico:

1. Existe una relación directa entre el rendimiento académico y el uso de TICs (redes sociales, juegos, navegación celular y videos), a mayor tiempo de exposición a plataformas como las redes sociales o videojuegos, especialmente dispositivos móviles, menor será el rendimiento académico.

2. Existe una relación positiva entre usar servicios de chat, compartir fotos y navegar en PC con el rendimiento académico, siempre y cuando los tiempos de exposición no superen las 3 horas por día.

En estos resultados nuevamente se encuentra rasgos positivos y negativos en el uso de redes sociales y el rendimiento académico, pero todo parece apuntar a que, si el tiempo de uso de dichas redes es moderado, el estudiante se enfoca en temas académicos y de intercambio de información y el profesor promueve actividades participativas entre los alumnos el rendimiento académico tiende a mejorar, mientras que si se usa excesivamente las redes sociales con fines no académicos; es decir, solo buscando el ocio, el rendimiento del alumno tenderá a disminuir.

2.4. Minería de datos.

Desde tiempos atrás, los avances tecnológicos han ido evolucionando cada día, facilitándonos el manejo y acceso de grandes cantidades de información almacenadas en un repositorio o base de datos. Esta información no es suficiente analizarla tan solo con herramientas estadísticas o las tradicionales de gestión de datos, sino que es necesario emplear técnicas de análisis de datos para grandes volúmenes de datos, la que ayudará a la extracción de información relevante para la futura toma de decisiones, este proceso es el de Minería de datos.

Desde el punto de vista de Molina y García (2006) dicen que “la Minería de Datos o Data Mining es un término genérico que engloba resultados de investigación, técnicas y herramientas usadas para extraer información útil de grandes bases de datos” (p. 5).

Según Angeles y Santillán (1988) expresan que la minería de datos es:

El proceso que tiene como propósito descubrir, extraer y almacenar información relevante de amplias bases de datos, a través de programas de búsqueda e identificación de patrones y relaciones globales, tendencias, desviaciones y otros indicadores aparentemente caóticos que tienen una explicación que pueden descubrirse mediante diversas técnicas de esta herramienta. (p. 79,80).

La minería de datos es un proceso sólido que permite realizar una búsqueda de contenido importante en grandes volúmenes de información o datos, con el fin de hacer uso de las

técnicas apropiadas para analizar los datos y comprender el modelo inferido, alcanzando soluciones a problemas determinados de una organización.

2.4.1. Tipos de datos.

La minería de datos puede ser empleada para cualquier tipo de información; sin embargo las técnicas de minería de datos se la utilizan de manera diferente para cada tipo de datos Hernández, J., Ramirez, J., y Ferri, C.(2004).

Dentro de la minería de datos, los tipos de datos que se consideran son los numéricos y los categóricos, datos como los enteros, reales, cadenas de texto, etc.; según el mismo autor existen otros tipos de datos como: base de datos espaciales, base de datos temporales, base de datos documentales, base de datos multimedia y la Word wide web.

- **Base de datos relacionales:** Según Hernández et al.(2004) una base de datos relacional es una colección de tablas o relaciones compuestas por atributos o columnas y de tuplas o filas. Cada columna, almacena información sobre una propiedad o atributo establecido de la relación y cada tupla posee una clave única puesto que representa un objeto y abarca un conjunto de valores para cada uno de los atributos de una relación. Esta base de datos tiene como característica principal la existencia de un esquema relacionado; en donde los datos son estructurados.
- **Bases de datos espaciales o geográficos:** Como su nombre lo indica, engloba datos asociados con el espacio físico (ubicación, dimensión, forma), donde incluye dos clases de datos: los espaciales (localización o gráficos) y lo nominales (atributos no gráficos); permitiendo asegurar patrones entre los datos siendo estos organizados y estructurados, los datos geográficos facilitan la visualización de los datos en forma de mapas. Como ejemplos tenemos: imágenes médicas, redes de transporte o tráfico, región montañosa, etc.
- **Bases de datos temporales:** “Las bases de datos temporales almacenan datos que incluyen muchos atributos relacionados con el tiempo o en que éste es muy relevante, Son utilizadas para encontrar características de la evolución de distintos valores en la base de datos”, esta base de datos contiene datos históricos y actuales. Obtenemos una base de datos histórica cuando insertamos información en la base de datos y en ésta no

se realiza modificaciones ni se eliminan los datos; mientras que si en la base de datos se actualiza o elimina los datos tendremos una base de datos instantánea.

- **Bases de datos documentales:** “Las base de datos documentales son adecuados para información no estructurada., disponen de todas las funciones necesarias de almacenamiento, indexación, consulta y modificación de información textual., necesarios en todo tratamiento documental” (Concorbado y Gómez, 1994); también pueden contener documentos estructurados como semi-estructurados.
- **Bases de datos multimedia:** La base de datos multimedia se caracteriza por almacenar imágenes, videos, texto y audio, resistiendo información de gran tamaño. Estas bases de datos requieren de técnicas de almacenamiento como de búsqueda o sistema de ficheros.
- **La Word Wide Web:** Este tipo de base de datos contiene una gran cantidad de información en la web en donde se puede extraer información importante y útil, es por ello que la minería web debe enfrentar muchas dificultades como determinar a qué páginas se debe acceder y como escoger la información apropiada para descubrir conocimiento.

2.4.2. Tipos de modelo.

Un modelo da soporte a las estrategias para la toma de decisiones mediante la aplicación de algoritmos para los datos; el mismo que está conformado por un conjunto de datos, estadísticas y patrones, estos son utilizados para crear una predicción o deducción de las relaciones que se pueden dar a los nuevos datos. Hernández et al. (2004) indica que “la minería de datos tiene como objetivo analizar los datos para extraer conocimiento. Este conocimiento puede ser en forma de relaciones, patrones o reglas inferidos de los datos precisamente desconocidos o de una descripción más concisa” (p.12). Existen dos tipos de modelos o métodos: El modelo predictivo y el modelo descriptivo.

2.4.2.1. Modelo predictivo.

El modelo predictivo también es denominado como algoritmo supervisado, el mismo que se encarga de “predecir el valor de un atributo (etiqueta) de un conjunto de datos, conocidos otros atributos (descriptivos); a partir de datos cuya etiqueta se conoce se induce una relación entre dicha etiqueta y otra serie de atributos” (Moreno, Miguel, García, y Polo, 1997) (p. 3). Según

Hernández et al. (2004), este modelo permite estimar valores desconocidos de las variables dependientes a partir de otras variables como las variables independientes (variables predictoras) de la base de datos. En la práctica este modelo es empleado principalmente para técnicas estadísticas.

En la minería de datos se encuentran varias tareas que deben considerarse como un tipo de problema que debe resolverse mediante un modelo o algoritmo; por lo que en el modelo predictivo se encuentran algunas de las tareas que deben ser resueltas por este modelo en la minería de datos. Según Osorio-Zuluaga (2009) estas tareas son:

- Clasificación (o discriminación)
- Clasificación suave
- Estimación de probabilidad de clasificación
- Categorización
- Preferencias o priorización
- Regresión

2.4.2.2. Modelo descriptivo (descubrimiento del conocimiento).

El modelo de descubrimiento del conocimiento llamado también algoritmos no supervisados permite “identificar patrones que explican o resumen los datos; es decir, sirven para explorar las propiedades de los datos examinados, no para predecir nuevos datos” (Hernández et al., 2004) (p. 12). Este modelo realiza tareas descriptivas como el fin de descubrir patrones y tendencias en los datos actuales, por lo que no utiliza datos históricos (Moreno et al., 1997). Este modelo describe el comportamiento de un conjunto de datos utilizando técnicas de la inteligencia artificial. Según Osorio-Zuluaga (2009) las tareas del modelo descriptivo son:

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

2.4.3. Proceso de extracción de conocimiento.

EL Knowledge Discovery in Databases (KDD) según Fayyad, Piatetsky-Shapiro, y Smyth, (1996) comentan que “es el proceso no trivial de identificación en los datos de patrones válidos, nuevos, potencialmente útiles, y finalmente comprensibles” (p. 41). Para Molina y García (2006) argumentan que el KDD “es el proceso completo de extracción de información, que se encarga además de la preparación de los datos y de la interpretación de los resultados obtenidos” (p. 1).

En si el KDD es un proceso de extracción de conocimiento a partir de una base de datos que contiene una sucesión repetidas de fases, dando como resultado un proceso iterativo, la cual conlleva a la elaboración de los modelos o patrones efectivos de los datos, evaluando e interpretando los resultados de los mismos (Figura 6). A continuación, se presentan las fases del KDD.



Figura 6. Proceso de KDD

Fuente:(Hernández et al., 2004) (p.13)

2.4.3.1. Fase de integración y recopilación de datos.

Formalmente César Pérez y Santín (2007) comentan que “las primeras fases del KDD determinan que las fases sucesivas sean capaces de extraer conocimiento válido y útil a partir de la información original” (p. 4). Así mismo, Sarasa, Suárez, y Sánchez (2008) indican que en esta fase “se determinan las fuentes de información que pueden ser útiles y dónde conseguirlas; se transforman todos los datos a un formato común, y se detectan y resuelven las inconsistencias” (p. 2).

Al hablar de integración de numerosas bases de datos nos referimos a la tecnología de almacenes de datos, la misma que trata sobre la tendencia actual de las empresas de recopilar datos de las bases de datos transaccionales y de varias fuentes para su análisis y por consiguiente para la toma de decisiones (Hernández et al., 2004).

2.4.3.2. *Fase de selección, limpieza y transformación.*

Después de haber realizado el primer paso de recopilar los datos, el siguiente paso a seguir es la selección de los datos que se va a minar, eliminando los datos irrelevantes que pueden afectar a la calidad de los datos. “En esta etapa, la confiabilidad en los datos se eleva. Incluye la claridad en los datos, tal como el manejo de valores faltantes y la remoción de ruido o datos anómalos” (Osorio, 2009).

Existen otros problemas que pueden afectar la calidad de los datos, como la presencia de datos anómalos que no se ajustan al comportamiento general de los datos llamados también outliers. Los algoritmos que cuentan la minería de datos no toman en cuenta a este tipo de datos, descartándolos. En esta etapa también se considera como problema los datos perdidos, ya que pueden ocasionar resultados no muy exactos; aunque existen procedimientos que pueden manejar este tipo de datos (Hernández et al., 2004).

En esta fase es importante eliminar la mayor cantidad de datos errores e irrelevantes, utilizando herramientas de consulta y estadística. Durante la selección de los datos se utilizan herramientas como histogramas y diagramas, con el fin de detectar datos inconsistentes o atípicos. En el proceso de transformación se utiliza técnicas de discretización que consiste en la reducción o el aumento de la dimensión de los valores, esta técnica permite transformar la información de los datos continuos en discretos (César Pérez y Santín, 2007).

2.4.3.3. *Fase de minería de datos.*

EL KDD tiene la característica primordial de todo el proceso, denominada fase de minería de datos; la cual consiste en ocasionar nuevos conocimientos o patrones que puedan ser utilizados por el usuario. Para lo cual se construye un modelo con el objetivo de realizar una descripción de los patrones y las relaciones que existen entre los datos explicando situaciones pasadas; es decir, explica predicciones que se pueden dar en los datos (Hernández et al., 2004).

Dentro de esta fase existen técnicas de minería de datos, las cuales se utilizan de acuerdo al tipo de conocimiento que se quiere extraer, considerando la técnica y el algoritmo más adecuado. En esta fase se concluye la tarea de la minería que se debe utilizar como la clasificación, regresión o agrupamiento.

2.4.3.4. Fase de evaluación e interpretación

Según Sarasa et al. (2008) “en la fase de evaluación e interpretación se evalúan los patrones y se analizan por expertos, y si es necesario, se vuelve a las fases anteriores para una nueva iteración” (p.2). En esta fase se pueden producir varias hipótesis de los modelos, cuyos modelos deben ser válidos e interpretables. Para que un modelo sea válido “es necesaria una fase consistente en el uso de criterios de evaluación de hipótesis. El despliegue del modelo a veces es trivial pero otras veces requiere un proceso de implementación o interpretación” (César Pérez y Santín, 2007) (p. 5).

Existen algunas técnicas de evaluación para probar un modelo, es por ello que se realiza una separación de los datos en dos conjuntos: de entrenamiento y de prueba o de test. Los métodos utilizados para la evaluación son: validación simple, validación cruzada con n pliegues y el bootstrapping (Hernández et al., 2004) .

2.4.4. Técnicas de minería de datos.

Desde el punto de vista de César Pérez y Santín (2007), “las técnicas de minería de datos persigue el descubrimiento automático del conocimiento contenido en la información almacenada de modo ordenado en grandes bases de datos” (p. 1). Para la extracción de conocimientos, es necesario seguir los pasos que conlleva la elaboración de un proyecto de minería de datos, independientemente de la técnica que se utilice (S. Vallejo, 2006). Es por ello que la minería de datos cuenta con varias técnicas que permiten extraer desde grandes volúmenes de datos el conocimiento de la información necesaria de acuerdo al enfoque del proyecto.

Desde el punto de vista de César Pérez y Santín (2007), las técnicas de minería de datos se clasifican en tres grupos:

- 1) Técnicas predictivas o supervisadas:** Se trata de métodos explicativos del análisis multivariante; en donde, las variables se clasifican en dependiente e independientes. a su vez se clasifican en regresión, análisis de la varianza y covarianza, series temporales, métodos bayesianos, algoritmos genéticos, y clasificación ad hoc (árboles de decisión, discriminante, redes neuronales).

- 2) Técnicas descriptivas o no supervisadas:** Son métodos descriptivos del análisis de multivariante, tienen la siguiente clasificación: clasificación post hoc (clustering, segmentación), asociación, dependencia, reducción de la dimensión, análisis exploratorio, escalamiento multidimensional.
- 3) Técnicas auxiliares:** dirigidas a la verificación, M. Pérez (2014) indica que “son herramientas más superficiales y limitadas. Son nuevos métodos basados en técnicas estadísticas descriptivas e informes clasificándose en proceso analítico de transacciones (OLAP) y reporting” (p. 2).

Las técnicas predictivas y las técnicas descriptivas se encuentran orientadas al descubrimiento del conocimiento concentrado en los datos. La Tabla 2, presenta la relación que existe entre las técnicas mencionadas en este apartado con las tareas de minería de datos, para la obtención relevante de información sobre el descubrimiento de patrones de un conjunto de datos.

Tabla 2. Relación entre las técnicas con las tareas de minería de datos

Técnica (algoritmo)	Predictivas		Descriptivas		
	Clasificación	Regresión	Agrupamiento	Asociación	Correlación
Redes Neuronales	X	X	X		
Árboles de decisión (ID.3, C4.5, C5.0)	X				
Árboles de decisión (CART)	X	X			
Árboles de decisión y sistemas de reglas (CN2)	X			X	
Reglas de Kohonen			X		
Modelización Estadística (Regresión lineal), (Regresión Logarítmica)		X			X
Modelización Estadística (Regresión Logística)	X			X	
Método basados en casos y en vecindad (k-means)			X		
Reglas de Asociación y Dependencia (A priori)				X	
Métodos Bayesianos (Naive Bayes)	X				
Métodos basados en casos y en vecindad (vecinos más próximos)	X	X	X		

Fuente: (Hernández et al., 2004)

A continuación se detallan algunas de las técnicas más representativas de la minería de datos son:

- **Redes neuronales:** Son técnicas no lineales que obtienen información a partir de grandes volúmenes de datos. Hernández et al. (2004) indican que “se trata de técnicas que aprenden un modelo mediante entrenamiento de los pesos que conectan un conjunto de nodos o neuronas. La topología de la red y los pesos de las conexiones determinan el patrón aprendido” (p. 147). En las tareas de reconocimiento de patrones y sistemas de clasificación se utiliza las redes neuronales, ya que la misma permite generalizar los patrones.
- **Regresión lineal:** La regresión lineal es un método estadístico predictivo que permite determinar la relación que existe entre variables, una dependiente y una o más variables dependientes, prediciendo un amplio número de sucesos. Es una técnica estadística común utilizada para resumir datos y realizar pronósticos ajustando una superficie o línea recta que minimice las discrepancias existentes entre los valores de salida reales y los pronosticados (D. Vallejo y Tenelanda, 2012). Es la más utilizada para formar relaciones entre datos, se caracteriza por ser rápida y eficaz aunque tiene la desventaja de ser insuficiente en áreas multidimensionales donde puedan asociarse más de 2 variables.
- **Árboles de decisión:** Los árboles de decisión están inmersos en los modelos predictivos o supervisados. Vizcaino (2008) menciona que “un árbol de decisión es un conjunto de condiciones o reglas organizadas en una estructura jerárquica, de tal manera que la decisión final se puede determinar siguiendo las condiciones que se cumplen desde la raíz hasta alguna de sus hojas” (p. 11). Según Hernández et al. (2004) los árboles de decisión o de clasificación se basan en dos tipos de algoritmos denominados divide y vencerás y los algoritmos denominados separa y vencerás. La minería de datos utiliza de manera más frecuente la técnica de los árboles de decisión por ser algoritmos clasificadores potenciales, además son algoritmos que tienen la facilidad de que sus resultados sean comprendidos por cualquier usuario.
- **Reglas de Asociación:** Conti y Martínez de Pinzón (2007) argumenta que “las reglas de asociación en la Minería de Datos (MD) representan una de las técnicas más valiosas e

interesantes para la extracción de conocimiento oculto en grandes volúmenes de datos” (p. 1). Las reglas de asociación son una técnica no supervisada de la minería de datos que permiten encontrar relaciones o correlaciones entre los elementos de una base de datos transaccional, obteniendo patrones de comportamiento entre los datos, estas reglas trabajan con atributos nominales. Hernández et al. (2004) menciona que “las reglas de asociación no implican una relación causa y efecto, es decir, puede no existir una causa para que los datos estén asociados” (p. 26). Estas reglas permiten una fácil interpretación de los datos por lo que no es necesario tener mucho conocimiento en minería de datos.

- **Método basados en casos y en vecindad (k-means):** Este método es el mas utilizado en cuanto al agrupamiento de los datos, por ser sencillo, veloz y eficaz. El algoritmo k-means “parte de un numero determinado de prototipos y un conjunto de ejemplo a agrupar sin etiquetar” (Ordóñez, 2013) (p. 17), la cual tiene como obeitivo ubicar a los centroides de manera que los datos que corresponden al mismo centroide adquieran características comunes para ello se utiliza la distancia euclidiana.

2.4.5. Reconocimiento de patrones

2.4.5.1. Factorización (Análisis Factorial).

El análisis factorial o factorización es un método estadístico multivariante lineal, cuyo propósito es reducir los datos que sirven para encontrar grupos homogéneos de variables a partir de un conjunto extenso de variables denominadas como factores (De la Fuente, 2011c).

Según Peña (2002) existe dos diferencias importantes entre el análisis factorial con los componentes principales, aunque los dos están relacionados, estas diferencias son:

- “Los componentes principales se construyen para explicar las varianzas, mientras que los factores se construyen para explicar las covarianzas o correlaciones entre las variables.
- Los componentes principales son una herramienta descriptiva, mientras que el análisis factorial presupone un modelo estadístico formal de generación de la muestra dada” (p. 137).

El análisis factorial simplifica la información dada en una matriz de correlaciones. En el análisis factorial todas las variables del análisis son independientes, por lo que no existe a priori una dependencia entre unas variables sobre otras (De la Fuente, 2011c). En el análisis factorial existen dos tipos de análisis: el exploratorio y el confirmatorio.

- **Análisis factorial exploratorio:** También denominado como técnica estadística de interdependencia; tiene las características de ser versátil por lo que todas sus variables son analizadas en conjunto (Martínez y Sepúlveda, 2012). Tiene como objetivo establecer la búsqueda por estructuras subyacentes desconocidas entre las variables de análisis, definiendo factores que se encuentren fuertemente correlacionados entre sí; “el hecho de que dos variables se encuentran relacionadas entre sí proviene del hecho que ambas variables poseen algo en común y algo que las diferencia” (E. Pérez y Medrano, 2010) (p. 56). Por ello, este análisis se lo utiliza para determinar o generar nuevos factores, ya que los mismos son desconocidos, a partir de un conjunto de variables, o también permitir reducir el número de estas (Díaz De Rada, 2002; Martínez y Sepúlveda, 2012).
- **Análisis factorial Confirmatorio:** “Permite contrastar un modelo construido con antelación, en el que el investigador establece a priori el conjunto total de las relaciones entre los elementos que lo configuran” (Herrero, 2010) (p. 1). En este tipo de análisis el investigador establece un nivel de confianza para poder determinar la verificación de hipótesis de manera concreta y específica (Martínez y Sepúlveda, 2012).

Según De la Fuente(2011b), “el análisis factorial se encarga de analizar la varianza común a todas las variables. Partiendo de una matriz de correlaciones, trata de simplificar la información que ofrece”.

Por tanto, en el Análisis Factorial para analizar toda la varianza se utiliza el método de análisis de componentes principales y para analizar solo la varianza común, se utiliza el análisis de factores comunes.

- **Análisis de componentes principales:** Es una técnica estadística de reducción de dimensión, que facilita el análisis e interpretación de los datos. Según Aldehuela (2005), “el análisis de componentes principales muestra la relación entre las distintas variables

de entrada, determinando su importancia y permitiendo reducir su número quedándonos únicamente con las más relevantes” (p. 95).

- **Análisis de factores principales:** Este método también se lo denomina como el método de ejes principales, y es un método iterativo en el que existe una reducción de la matriz de correlaciones original. Este análisis calcula únicamente en base a la varianza común los factores entre las variables y sustituye la correlación poblacional por la correlación muestral (Malhotra, 1997; De la Fuente, 2011b), estos valores son asignados a la matriz de correlaciones específicamente en su diagonal.

Según De la Fuente (2011b) el análisis factorial debe seguir los siguientes pasos:

- Formulación del Problema
- Análisis de la matriz de correlación
- Extracción de factores
- Determinación del número de factores
- Rotación de factores
- Validación del modelo: mediante el cálculo de puntuación factoriales o por la selección de las variables representativas
- Finalmente análisis posteriores utilizando otras técnicas como regresión, clústers, etc. (p. 4)

Matriz de correlaciones.

La matriz de correlación es una matriz construida por n filas y n columnas, dando como resultado una matriz simétrica; esta matriz está conformada por los coeficientes de correlación de cada par de variables. La matriz de correlaciones es la base para el análisis factorial de las variables, por lo que permite observar cómo se encuentran relacionadas cada uno de las variables con respecto a otras variables. La matriz de correlaciones es definida por el coeficiente de correlación de Pearson.

La matriz de correlación cuenta con varios indicadores para su análisis entre ellos tenemos:

- **Prueba de esfericidad de Bartlett:** Esta prueba se basa en verificar la hipótesis nula

que indica que la matriz de correlación es una matriz de identidad, lo que significa que las correlaciones entre las variables no son significativas, considerando que no es adecuado utilizar el análisis factorial; es decir, las variables que se encuentran en la diagonal principal de la matriz tienen el valor de uno y las demás variables tienen el valor de cero. Según Torres (2012), este test “utiliza una estimación de chi-cuadrado sobre el determinante de la matriz de correlaciones lo que permite aceptar o rechazar la hipótesis nula con nivel de significancia menor a 0,05” (p. 91)

- **Prueba KMO (Kaiser-Meyer-Olkin):** Quintín y Paz (2007) indica que “es un índice que permite comparar las magnitudes de los coeficientes de correlación observados con las magnitudes de los coeficientes de correlación parcial” (p. 328). El KMO toma los valores entre 0 y 1, de manera que cuanto más bajo sea su valor, el valor de los coeficientes de correlación parciales será mayor, por lo que el análisis factorial no será adecuado para utilizarlo (De la Fuente, 2011c); es decir, mientras el valor del KMO es menor a 0,5 el análisis factorial es inaceptable, caso contrario, si los valores son cercanos a 1 muestra que existe una correlación alta por lo que el análisis factorial es adecuado.
- **Métodos de rotación:** Permite medir el grado y tipo de la relación que existe entre los factores y las variables originales para facilitar la interpretación de los datos, la interpretación de los factores se facilita si los que afectan a algunas variables no lo hacen a otras y al revés (Morales, 2013). Entre los métodos de rotación tenemos:
 - Rotación Varimax: La rotación varimax es el procedimiento más utilizado, este método “maximiza la varianza de los factores, cada columna de la matriz factorial rotada tendrá cargas factoriales altas con algunas caríbeles y bajas con otras, lo cual facilita la interpretación” (Álvarez, 2007) (p. 237). La rotación varimax comprende valores de las cargas que sean cercanas al -1 o al +1 y otras cargas que sean cercanas al 0.
 - Rotación Cuartimax: Con la rotación Cuartimax “se intenta simplificar las filas , es decir se maximiza la varianza de las filas” (Morales, 2013) (p. 13). Según el punto de vista de R. Álvarez (2007) la rotación Cuartimax tiene como finalidad “simplificar las filas de la matriz factorial, de esta manera, que cada variable tendrá una correlación alta con pocos factores y baja con los demás, lo cual

facilitara la interpretación” (p. 237).

➤ Rotación Equamax: Este tipo de rotación es la combinación entre la rotación varimax, y la rotación Cuartimax. Esta rotación tienen como objetivo “minimizar tanto el número de variables que saturan alto un factor como el número de factores necesarios para explicar una variable” (César Pérez y Santín, 2007) (p. 361).

- **Criterio del porcentaje de varianza:** Este método consiste en “tomar como número de factores el número mínimo necesario para que el porcentaje acumulado de la varianza explicado alcance un nivel satisfactorio (75%, 80%)” (De la Fuente, 2011c) (p. 13), si el porcentaje es menor al señalado entonces no es adecuado el análisis factorial.

2.4.5.2. Clusterización (Análisis clúster).

Desde el punto de vista de Sierra-Araujo (2006) “el análisis de clúster o clustering, es una colección de métodos estadísticos que permiten agrupar casos sobre los cuales se miden diferentes variables o características” (p. 36). El análisis de clúster tiene por objetivo clasificar los elementos o variables de la muestra obtenida en razón a sus características similares, logrando una semejanza o diferencia en cada grupo.

El análisis clúster llamado también conglomerado de pertenencia suele iniciarse valorando las características comunes mediante la semejanza de las distintas variables ya sean cuantitativas o cualitativas. Posteriormente se procede a comparar los grupos según las similitudes, y finalmente se decide cuántos grupos se construyen; el objetivo será formar el mínimo número de grupos posible, lo más homogéneos dentro de cada grupo, y lo más heterogéneos entre grupos (Vilà, Rubio, Berlanga, y Torrado, 2014). Cuando se clasifica las variables se hace una división entre ellas y se forma grupos, permitiendo hacer una reducción en la dimensión de las variables.

En los métodos de clasificación juega un rol importante el índice de similitud o distancia utilizada pues de ello depende en gran medida que los resultados finales tengan la mayor confiabilidad posible Miranda (1987, citado por Chávez, Miranda, Varela, y Fernández, 2010). El análisis de clúster es un método que permite descubrir asociaciones y estructuras en los datos que no son evidentes a priori pero que pueden ser útiles una vez que se han encontrado (Vicente, 2011).

Por tal razón, en el análisis discriminante o en la regresión logística se conocen los grupos que se han clasificado con mayor exactitud.

Según Pérez (2009), los principios básicos del análisis de conglomerados son:

- Es un método estadístico multivariante de clasificación automática de datos.
- Tiene la finalidad de revelar concentraciones en los datos para un agrupamiento eficiente en conglomerados según su homogeneidad.
- El agrupamiento puede ser para casos o variables (cualitativas o cuantitativas).
- Es esencial un uso adecuado del concepto de distancia, ya que los grupos se realizan según la proximidad o lejanía de unos con otros.
- Es fundamental que dentro de un conglomerado los elementos sean homogéneos, y lo más diferentes a los contenidos del resto.
- Es una técnica de clasificación post hoc: el número de conglomerados se determina en función de los datos, y puede no ser definido de antemano (p. 28).

El análisis clúster se divide en dos métodos: el jerárquico y no jerárquicos o particionales, cuya diferencia principal entre los mismos es la perspectiva o visión que se le brinda al proceso de asignación de los grupos a los clúster.

Método Jerárquico.

La clasificación jerárquica inicia con tantos clúster como observaciones y termina en un solo clúster que contiene todas las observaciones (Rencher, 2002), uniéndose al final en un solo grupo. Este método es más apropiado utilizarlo para muestras pequeñas. Estos métodos según Sierra-Araujo (2006) se dividen en:

- Métodos aglomerativos: Se caracterizan por ser rápidos y crear clústeres agrupándolos; es decir, se unen los clústeres de dos en dos hasta que crean un único clúster.
- Métodos divisivos: Este método a diferencia de los métodos aglomerativos son lentos por lo que se los utiliza para datos con pocos casos. Estos se caracterizan por crear clúster dividiéndolos., es decir comienza de un único clúster, se divide en grupos más pequeños hasta llegar a que cada caso constituya un único clúster. (p.125)

Los métodos jerárquicos son representados por dendogramas, mostrando el proceso de unión o división de los clústers. Dentro de estos métodos tenemos: Algoritmo básico de clasificación, algoritmo de clasificación, el método del mínimo (vecino más próximo), el método del máximo (vecino más lejano), el método de Ward, formula de Lance-Williams y clúster basado en distancias.

Métodos Particionales.

La clasificación no jerárquica o particionales es un método adecuado para muestras más grandes, en donde, como entrada se tiene un número de grupos que son asignados a grupos diferenciados sin depender unos con otros. Cuando el número de observaciones de la muestra es alto se recomienda trabajar con métodos no jerárquicos (Díaz De Rada, 2002), caracterizándose por ser sencillo y eficiente.

Los métodos particionales también se los denomina métodos de optimización, y a diferencia de los métodos jerárquicos, estos utilizan una matriz de distancias o similitudes. En los métodos particionales “se debe fijar de antemano el número de clústers k que tendrá la partición”. (Sierra-Araujo, 2006) (p. 127), por lo que se pueden utilizar varias técnicas para llegar a estimar ese valor.

Desde el punto de vista de R. Álvarez (2007) “los métodos particionales forman el número de grupos indicado por el usuario, de tal manera que la varianza dentro de cada grupo sea mínima” (p. 221). Estos métodos dan lugar a una partición mediante la minimización o maximización de algún criterio; es por ello que depende del criterio de optimización y de los algoritmos empleados para obtener la partición de los clústers (Sierra-Araujo, 2006). Dentro de los métodos particionales se encuentra el algoritmo k -medias.

El Algoritmo k -medias.

Es uno de los algoritmos más utilizados que disminuye la suma de las distancias al cuadrado de cada modelo al centroide de su grupo, este método requiere que previamente se especifique el número de grupos que se desea extraer y parten de un conjunto N de semillas que luego son reemplazadas por los centroides de los clúster, tomándose como clúster cada una de las semillas empleadas (Rencher, 2002). El algoritmo de k -medias se caracteriza por la distancia

empleada en el análisis clúster denominada como euclídea, la misma que se usa como una métrica, como también, una de sus características importantes es la varianza que es empleada como la medida de la dispersión de los conglomerados.

De acuerdo a la clasificación k-medias propuesto por Xu, R. y Wunsch (2009), debe tener los siguientes pasos:

- 1) Inicializar cada partición K de forma randómica o en base a centroides preexistentes.
Calcular la matriz prototipo del clúster $M = [m_1 , \dots , m_K]$
- 2) Asignar cada sujeto al clúster más cercano
- 3) Recalcular la matriz prototipo del clúster en base a la partición actual
- 4) Repetir los pasos 2 y 3 hasta que no haya cambios en el centroide de los clúster (p. 144).

En el análisis no jerárquico no es necesario medir las distancias o semejanzas entre grupos, por tanto su categorización en k conglomerados se la realiza en un solo paso. Otro punto a tener en cuenta es que cuando se realice el análisis de conglomerados (clústeres) se debe acompañar los valores resultantes que pertenecen a uno u otro grupo con una Tabla descriptiva, resumiendo los valores que muestren las variables contenidas en cada grupo, permitiendo así determinar la característica principal en cada grupo, para su respectiva interpretación y proporcionando un medio de evaluación de cada conglomerado que es objeto de estudio (De la Fuente, 2011a).

Para evaluar la clasificación resultante en el análisis clúster es necesario utilizar otras técnicas que permitan validar los clústers; las técnicas más utilizadas son: el análisis de varianza y el análisis discriminante. El análisis de varianza permite conocer las diferencias de los factores en cada uno de los clústers, por otro lado el análisis discriminante permite obtener el porcentaje de sujetos que están determinados correctamente (Díaz De Rada, 1998).

2.4.5.3. *Análisis discriminante (clasificación).*

El análisis discriminante es un método predictivo de clasificación ad-hoc, inmerso en el análisis estadístico multivariante, que tiene por objetivo analizar las diferencias significativas que pueden existir entre los grupos o factores analizados en relación a un conjunto de variables,

permitiendo una clasificación sistemática (De la Fuente, 2011b); en este método el analista debe conocer de antemano la clasificación de los grupos mediante las diferencias que existen entre los mismos, este método abarca las técnicas de clasificación supervisada.

Sierra-Araujo (2006) menciona que existen tres objetivos fundamentales planteados en el análisis discriminante:

- Proporciona un mejor entendimiento de las diferencias existentes entre los grupos analizados con respecto a las variables medidas en el estudio (análisis discriminante descriptivo)
- Construir procedimientos sistemáticos de clasificación de nuevas observaciones en dichos grupos que gocen de buenas propiedades en termino de probabilidades de error (análisis discriminante predictivo)
- Determinar la forma más parsimoniosa de distinguir entre los grupos descartando aquellas variables clasificadoras que estén poco relacionadas con las diferencias existentes entre los grupos (algoritmos de selección de variables) (p.208).

Es muy importante acotar que para el análisis discriminante se deben tener en cuenta los siguientes supuestos según De la Fuente (2011b).

- La distribución de las variables independientes deben estar bajo la hipótesis de la normalidad.
- No debe existir una combinación lineal entre las variables discriminantes con otras variables discriminantes.
- Las variables de razón o intervalo son independientes de la variable categórica.
- Debe existir al menos dos grupos y en cada grupo constan dos o más casos.
- En cada grupo las matrices de covarianzas deben de ser aproximadamente similares.
- Debe existir una distribución normal multivariante en las variables continuas (p.1).

El mismo autor comenta que este método es considerado como “un análisis de regresión donde la variable dependiente es categórica y tiene como categorías la etiqueta de cada uno de los grupos, mientras que las variables independientes son continuas y determinan a qué grupos pertenecen los objetos” (p. 1).

2.4.5.4. Regresiones.

La regresión es una tarea predictiva que consiste en aprender una función real que asigna a cada instancia un valor real, cuyo valor a predecir es numérico, permitiendo minimizar el error entre el valor predicho y el valor real (Hernández et al., 2004). En la regresión las variables de respuesta o dependientes y las variables explicativas o independientes son cuantitativas.

Regresión Logística.

Vieira, Ortiz, y Ramirez (2009) indica que “el modelo de regresión logística ha sido utilizado para resolver problemas de clasificación supervisada” (p. 45), teniendo como objetivo pronosticar el riesgo o probabilidad de que ocurra o no un suceso en un tiempo específico, como una función de las variables explicativas.

La regresión logística se utiliza para estudiar el efecto de múltiples variables explicativas sobre una variable de respuesta categórica, ya sea dicotómica o más de dos categorías en escala ordinal. Taucher (1997) menciona que “en estos casos no se puede aplicar el modelo de regresión lineal múltiple por no cumplirse el supuesto de continuidad y distribución normal de la variable dependiente” (p. 241).

Esta regresión tiene una ligera diferencia con la regresión lineal múltiple, con respecto a sus enfoques matemáticos, además la regresión logística se caracteriza porque su variable dependiente es métrica y discreta, no es continua y su variable de interés o variable independiente es una variable dicotómica.

Los modelos de regresión logística son muy atractivos por su utilidad en el cálculo de probabilidades, pero hay que tener en cuenta que es una técnica compleja, por lo que hay que tener especiales precauciones para no cometer errores importantes (Álvarez, 2007).

La regresión logística identifica el mejor modelo a través del cociente de verosimilitud, el mismo que determina si una variable debería eliminarse del modelo (Hernández-Rodríguez, 1998), e indica cuanto es más probable un modelo con respecto a otro. “La diferencia de los cocientes de verosimilitud entre dos modelos se distribuye según la ley de la Chi-cuadrado con los grados de libertad correspondientes a la diferencia en el número de variables entre ambos modelos”

(De la Fuente, 2011d) (p. 1).

- **Regresiones logística binomial:** La regresión logística binomial es muy importante en el área de la medicina ya que su modelo se adapta a esta área. La variable dependiente es una variable dicotómica que tomara los valores 0 ó 1; por ejemplo éxito o fracaso, ausencia o presencia, hombre o mujer, etc. Según De la Fuente (2011d) “el objetivo de la regresión logística binomial es hallar los coeficientes que mejor se ajusten a la expresión funcional” (p. 5).
- **Regresiones logística multinomial:** La regresión logística multinomial se utiliza para modelos que tienen la variable dependiente de tipo nominal con más de dos categorías, es decir que su variable sea politómica (Hosmer y Lemeshow, 2004). Las variables dependientes o explicativas son categóricas o cuantitativas.

El modelo de regresión logística cuenta con varios test estadísticos que ayudan a verificar la robustez del modelo (De la Fuente, 2011d). Estos son:

- **Odds Ratio:** El OR se obtiene de la división de dos Odds, es decir, del cociente entre dos probabilidades: el grupo de interés y el grupo de referencia. El OR permite identificar la fuerza de la relación entre las dos probabilidades en estudio, lo que ayuda hacer comparaciones entre ellas. En SPSS se lo representa Exp (b) (Torres, 2012).
- **Bondad de ajuste:** La bondad de ajuste permite comparar el ajuste del modelo producido o resultante con el modelo evaluado, ayudando a concluir si este modelo puede ser recomendable para el estudio, donde su desviación debe ser igual o menor que sus grados de libertad. “Se considera que un modelo no presenta un buen ajuste si la variabilidad residual es grande, sistemática o no se corresponde con la variabilidad descrita por el modelo” (Ryan, 1997) (p. 335). Para poder medir la adecuación de un modelo es necesario contar con medidas generales de bondad de ajuste.
- **Logaritmo de la verosimilitud:** El logaritmo de la verosimilitud (-2 Log likelihood), mide hasta qué punto un modelo se ajusta bien a los datos. Este test El resultado de esta medición recibe también el nombre de desviación. Cuanto más pequeño sea el valor, mejor será el ajuste (De la Fuente, 2011d). Este test tiene como objetivo mostrar la

estimación que otorgue una máxima probabilidad a los datos, con la finalidad de realizar una comparación de un modelo con respecto a otro, es decir que verifica la hipótesis nula, la misma que indica que los coeficientes de las variables son igual a cero, mientras que la hipótesis alternativa dice que al menos un coeficiente de las variables es diferente de cero. Este test presenta una significancia cuando el valor del chi-cuadrado es ($p < 0,05$), descartando así la hipótesis nula. Cabe destacar que el resultado que expresa el logaritmo de la verosimilitud es la resta de los valores del modelo de intercepción y el modelo ajustado final (De la Fuente, 2011d).

- **Pseudo R2:** Este indicador expresa la variabilidad de la variable dependiente que es explicada por el modelo. Los valores próximos a 1 serán indicadores de muy alto ajuste y capacidad predictiva, mientras que valores cercanos a cero es todo lo contrario (Canela, Lora, y Estrella, 2011). La R cuadrado de Nagelkerke es otro de los estadísticos de la bondad de ajuste, por lo que es una versión mejorada de la R cuadrado de Cox y Snell, toma el valor correspondiente al porcentaje de varianza explicada de los datos reales recogidos por el modelo (De la Fuente, 2011d).
- **Test de Wald:** El test de Wald es un estadístico que permite contrastar la hipótesis nula que indica que su coeficiente de cada variable independiente es igual a cero, en cambio la hipótesis alternativa contrasta que los coeficientes de cada variable son distintos de cero, concluyendo que existe una asociación muy significativa entra la variable dependiente e independiente (Torres, 2012). El test de Wald y la razón de verosimilitud tienen algunas similitudes; sin embargo, se obtienen resultados diferentes, por lo que el test de Wald maneja dos aproximaciones; en cambio, la razón de verosimilitud utiliza una aproximación. Hernández-Rodríguez (1998) indica que “el logaritmo de Wald no es confiable cuando el coeficiente de regresión es muy grande” (p.164).

CAPÍTULO 3: METODOLOGÍA

En este capítulo se presenta la metodología empleada para llegar a obtener los objetivos esperados en la investigación del uso del Internet en las universidades ecuatorianas, en este caso, se realizó el estudio en la UNL, universidad pública ecuatoriana de tipo B, que consta de 5 áreas académicas y 34 carreras.

3.1. Población y tamaño de la muestra.

La población universitaria investigada corresponde a estudiantes que realizan sus estudios en la modalidad presencial, donde se estima alrededor de 8318 estudiantes matriculados en el periodo académico Octubre 2014 - Febrero 2015 (Anexo 1), que es indispensable para el cálculo del tamaño de la muestra en poblaciones finitas. En esta investigación se ha determinado que el nivel de confianza o credibilidad que se asignarán a los resultados obtenidos es del 95% con un margen de error del 5%.

Arnal, del Rincón, y Latorre (1992, citados por P. Muñoz y González, 2010) (p. 5), quienes proponen una fórmula para calcular el tamaño de la muestra para poblaciones finitas:

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

N= Total de la población

Z²= 1,962 (seguridad del 95%)

p= probabilidad de éxito 0.50

q= probabilidad de fracaso 0.50

d = Error de estimación 5% = 0.05

$$n = \frac{8318(1,96)^2 0.50 * 0.50}{(0.05)^2(8318 - 1) + (1,96)^2 0.50 * 0.50} = 367.24 \cong 368$$

Al aplicar la fórmula para determinar el tamaño de la muestra, da como resultado 368 personas que se debe encuestar como mínimo para que la investigación sea factible; no obstante, para el estudio investigativo se han encuestado a 504 estudiantes, de las diferentes áreas (<https://www.youtube.com/watch?v=VIQjY5tmhCI>), a quienes se les pidió colaboración en la aplicación a la encuesta elaborada por el estudio (Anexo 2) y la entrevista (<https://www.youtube.com/watch?v=VIQjY5tmhCI>) aplicada al docente del área de Sistemas

Informáticos de la universidad en estudio (Ing. Roberto Jácome), la misma que consta de temas fundamentales: Instalaciones físicas y políticas de uso de tecnología.

En el primer tema se recoge información sobre las salas de cómputo, Internet inalámbrico, ancho de banda, número de estudiantes, software para la gestión académica y administrativa, campus virtual, matriculación en línea, consultas de calificación en línea, acceso a biblioteca virtual y porcentaje de profesores que cuentan con computador personal. El segundo tema comprende: la interacción de los docentes con los elementos tecnológicos para sus actividades de enseñanza, plan de formación de profesores en el uso de las tecnologías para la educación, porcentaje de profesores capacitados en temas tecnológicos y pedagógicos y si existe incentivos para los profesores que utilizan las TIC en sus actividades académicas.

3.2. Análisis e interpretación de los datos.

La metodología utilizada en el desarrollo de la presente investigación es analizada con herramientas estadísticas y empleando principalmente el proceso de minería de datos como es el proceso de Knowledge Discovery in Databases (KDD) que según Hernández et al. (2004) consta de las siguientes fases: “(1) fase de integración y recopilación de datos, (2) fase de selección, limpieza y transformación de los datos, (3) fase de minería de datos, (4) fase de evaluación e interpretación” (p. 20). Figura 7.

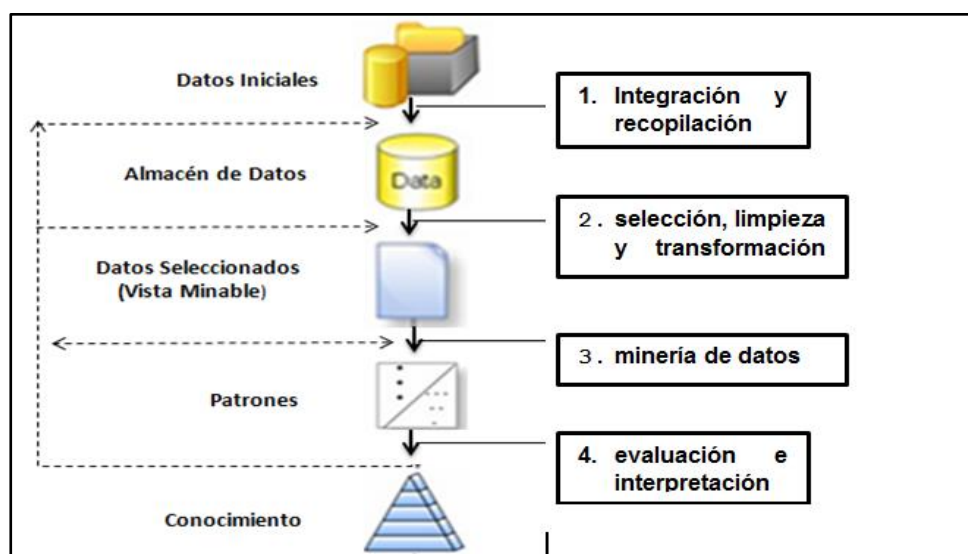


Figura 7. Fases del KDD

Fuente:(Hernández et al., 2004) (p. 20)

A continuación, se describe el proceso proporcionando información indispensable para alcanzar los objetivos deseados en este estudio; así mismo, permitirá dar a conocer la confiabilidad de los resultados. Es necesario centrarse en explicar cada uno de los métodos empleados para dar contestación a las hipótesis planteadas en la investigación.

3.2.1. Fase de integración y recopilación de datos.

La fase de integración y recopilación de datos es la primera fase en el proceso de extracción de conocimiento de una base de datos, integrándolos en este caso en la Dataset del proyecto de investigación, teniendo como resultado un almacén de datos que serán manejables de acuerdo a las etapas del proceso KDD.

Para realizar el levantamiento de información en la UNL como tema de estudio, se realizó una carta dirigido al Rector de la universidad (Anexo 3), con la finalidad de obtener el permiso respectivo para el acceso a cada una de las áreas y así poder obtener los datos que permitan llegar a los objetivos propuestos para la investigación. Gracias a la favorable y exitosa contestación al pedido realizado por parte de la universidad se tuvo el acceso a la información solicitada en donde fueron participes los estudiantes, como también, uno de los docentes que forma parte del área de Sistemas informáticos (Anexo 4).

Se consideró un conjunto de estrategias para obtener los resultados esperados. Como instrumento de recolección de datos se utilizó la técnica de la encuesta y la entrevista. La encuesta empleada para el proyecto de uso de las tecnologías de Internet en las universidades del Ecuador es proporcionada por el Director del Proyecto, quién se basó en los “proyectos PIC, DLINHE, Ecuador, apropiada al entorno de los estudiantes de la modalidad presencial de las Universidades de tipo A, B, C y a las necesidades del presente estudio” (Anexo 20). La encuesta consta de 17 ítems (Anexo 2) información se recogió manualmente y se examinó lo siguiente:

- Datos generales de la población universitaria: comprenden las primeras preguntas de la encuesta, recogida a través de las variables: carrera, edad, género y nivel de ingresos; así mismo, se considera aspectos socio-económicos de los estudiantes y generales.
- Información sobre el uso del Internet en las preferencias del estudiante. (preguntas 6 a la 9, con las variables: lugar de conexión, medida en 5 niveles (1: desde la casa, 2: desde un cyber café, 3: desde el trabajo, 4: desde la universidad, 5: desde una red móvil)

(variables categórica), conexión diaria de Internet, con 7 niveles, nivel de conocimiento medida en 10 niveles, horas que se conecta al día, los años de experiencia de Internet, además, se recogió información de las 3 variables agrupadas en la pregunta 13 que se refiere a las cuentas de blog, youtube y del.ici.us(variable categórica-dicotómica), como también de las 7 variables agrupadas en la pregunta 14 que consta de 10 niveles que habla acerca del nivel de uso de los dispositivos tecnológicos. (Anexo 2).

- Uso del Internet en las actividades académicas. 10 ítems que se reúnen en la pregunta 10 del Anexo 2.
- Uso del Internet para entretenimiento: información recogida a través de 5 variables concentradas en la pregunta número 11; así mismo, se levantó información sobre las cuentas creadas por los estudiantes en las tres variables agrupadas a la pregunta número 12 (Anexo 2).
- Generalidades sobre el rendimiento académico. Esta información es recogida a través de 2 ítems agrupadas de la pregunta 17 (Anexo 2).
- La pregunta 15 comprenden las percepciones de los estudiantes acerca de la importancia del uso del Internet para realizar sus trabajos académicos, consta de 6 variables medida con 10 niveles; de igual forma la pregunta 16, recoge la información en 11 variables de tres niveles (1: si, 2: no, 3: a veces), referente a las percepciones de los estudiantes con respecto al uso de las herramientas tecnológicas por parte de los profesores.

Tabuladas las encuestas manualmente, se utilizó la herramienta en línea SurveyMonkey con la finalidad de subir la información necesaria para la creación de una base de datos preliminar (DataSet) que ayudará a desarrollar los análisis estadísticos para la solución de las hipótesis plantadas en la investigación.

3.2.2. Fase de selección, limpieza y transformación de datos.

Luego de haber realizado el procedimiento de recolección de datos es importante seguir con la siguiente fase que indica el proceso KDD: selección, limpieza y transformación de datos, con el fin de obtener una vista minable de los datos más relevantes para la investigación, eliminando los datos erróneos o inconsistentes encontrados en la base de datos (DataSet). Esta fase permite seleccionar un subconjunto de datos que se va a minar, mejorando la calidad de los datos y proporcionando un subconjunto de datos más apropiado para resolver las interrogantes planteadas en la investigación.

- **Limpieza de datos:** Esta tarea se encarga de eliminar los datos anómalos o valores atípicos, registros con datos faltantes o irrelevantes, valores que causaron ruido, registros de datos en blanco y registros de estudiantes inconsistentes, por ejemplo: estudiantes que no se matricularon en el semestre anterior pero que registraban una respuesta en las materias aprobadas el semestre anterior. Estos datos pueden dar como resultados un análisis inexacto o incorrecto, por lo que es necesario la utilización de histogramas que permitan observar la distribución normal de los datos, eliminando aquellos que no favorecen en la investigación.
- **Transformación de datos:** Aquí se realizó la transformación de los atributos o variables, que consiste en tomar las variables existentes para generar nuevas variables que contengan una estructura más adecuada. Por tal motivo, las variables que constan en la transformación son:

Tabla 3. Variables inmersas en la transformación de los datos

Variable	Descripción	Valor origen	Valor de la Variable transformada	Tipo de dato	Medida
sex	Género del estudiante	Hombre Mujer	1:Hombre; 2: Mujer	Numérico	Dicotómica
ing	Ingresos mensuales de la familia del estudiante	Hasta 350; Hasta 600; Hasta 1000; Hasta 1500; Más de 1500 dólares	1: Hasta 350; 2: Hasta 600; 3: Hasta 1000; 4: Hasta 1500; 5: Más de 1500 dólares	Numérico	Nominal
lug_con	Lugar desde donde el estudiante se conecta a Internet habitualmente	Desde la casa; Desde un cyber café Desde el trabajo Desde la universidad Desde una red móvil	1: Desde la casa; 2: Desde un cyber café 3: Desde el trabajo 4: Desde la universidad 5: Desde una red móvil	Numérico	Nominal
rendimiento_academico	Módulo aprobado o no aprobado por el estudiante	Aprobó No aprobó	0: Aprobó 1: No aprobó	numérico	Dicotómica

Fuente: Elaboración propia

Como se puede observar se utilizó las operaciones de discretización y numerización con el propósito de derivar nuevas variables y cambiar el tipo de datos ya sea de un valor numérico a un valor nominal o viceversa. Así mismo, se realizó la normalización de los datos con el objetivo de consolidarlos de manera que puedan ser manejables para la siguiente fase del proceso KDD.

- **Selección de los datos:** En esta fase se identifica las variables relevantes que aportan información importante para la demostración de las hipótesis del proyecto. A continuación en la tabla 4, se detalla las variables seleccionadas.

Tabla 4. Variables seleccionadas para la comprobación de hipótesis

Hipótesis	Variables seleccionadas	Variables para la clusterización	Variable construida	Variables independiente	Variables dependiente
H1 y H3	ing a_ing_pla a_con_pro a_con_est a_rec_edu a_vid_aca a_for_vir a_post_aca a_hor_cha a_bus_inf a_bib_vir rendimiento_academico	a_con_pro, a_con_est, a_rec_edu a_vid_aca, a_bus_inf	Perfil para actividades académicas "cluster3_aca"	H1:ing H3: perfil para actividades académicas	H1: perfil para actividades académicas H3:rendimiento_academico
H2 y H4	ing e_hor_cha e_red_soc e_jue_lin e_des_con e_vid_sem rendimiento_academico	e_hor_cha, e_red_soc, e_vid_sem	Perfil para entretenimiento "cluster3_entre"	H2: ing H4: perfil para entretenimiento	H2: perfil para entretenimiento H4:rendimiento_academico

Fuente: Elaboración propia

3.2.3. Fase de minería de datos.

La fase de minería de datos es utilizada para definir todo el proceso como tal, su objetivo principal es la extracción del conocimiento que es utilizado por el usuario, de manera que permita construir un modelo que conste de patrones y relaciones entre los datos alojados en el DataSet; obteniéndose predicciones que ayuden a explicar mejor los datos. Por tal razón, es preciso seguir las siguientes medidas con el fin de obtener los mejores resultados (Hasperué, 2012).

3.2.3.1. Escoger el modelo más adecuado.

Para seleccionar el modelo que mejor se adapte al desarrollo de la investigación es necesario analizar los objetivos a los que se desea llegar como también dar contestación a las hipótesis planteadas al inicio del proyecto investigativo; por lo que se ha considerado seleccionar los dos tipos de modelos de la minería de datos: los descriptivos y los predictivos.

Los modelos descriptivos permitirán el agrupamiento de los estudiantes en grupos homogéneos; y los modelos predictivos ayudaran principalmente a la comprobación de las hipótesis, puesto que predicen la ocurrencia de una variable dependiente en función de una o más variables independientes.

3.2.3.2. Establecer la tarea de minería de datos.

En el presente estudio se consideraron las siguientes tareas de minería de datos con la finalidad de obtener la tarea más apropiada de clasificación y agrupamiento de los datos almacenados en el Dataset.

Factorización.

En esta investigación en primera instancia se utilizó una de las tareas de minería de datos denominada factorización denominada en estadística como análisis factorial (técnica multivariante), la que permite reducir las variables con el fin de encontrar grupos homogéneos sin perder información, esta técnica facilita el manejo de las nuevas variables denominadas en estadística como factores. Técnica que fue aplicada tanto en las actividades académicas como de entretenimiento. En donde se escogieron aquellas variables que se caracterizan principalmente por correlacionarse entre sí y por ser independientes. Sin embargo, se pudo evidenciar que las mismas no cumplieron con las condiciones que plantea la factorización tanto en las actividades académicas como de entretenimiento, cuyas condiciones según De la Fuente (2011b) 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.

Por esta razón, se descarta el uso de esta técnica y se realiza la experimentación con otras tareas de minería de datos como la clusterización (Anexo 5, 6).

Clusterización.

El análisis clúster llamado también conglomerado de pertenencia es otro método estadístico multivariante; dentro de la perspectiva de minería de datos esta tarea es denominada de agrupamiento o segmentación (clustering). Una tarea descriptiva y tiene como propósito la obtención de grupos con características similares, obteniendo como característica principal los grupos como tal y la determinación de la pertenencia de los mismos (Hernández et al., 2004). Las variables que tuvieron un tratamiento especial con esta tarea fueron las variables dependientes: académicos y de entretenimiento, los cuales constan de 10 y 5 variables correspondientemente. Se debe seleccionar un algoritmo que permita resolver la tarea escogida para obtener el modelo que se busca.

3.2.3.3. Elegir el algoritmo para minería de datos.

Los algoritmos que se encuentran dentro de esta tarea se clasifican en algoritmos paramétricos (jerárquicos) y no paramétricos denominados también como k-medias. Los algoritmos paramétricos inician con tantos clúster como observaciones y termina en un solo clúster que contiene todas las observaciones, uniéndose al final en un solo grupo (Rencher, 2002). Este método es más apropiado utilizarlo para muestras pequeñas. Por otro lado, los algoritmos no paramétricos son adecuados para muestras más grandes, en donde, como entrada se tiene un número de grupos que son asignados a grupos diferenciados sin depender unos de otros. Cuando el número de observaciones de la muestra es alto se recomienda trabajar con métodos no jerárquicos (Díaz De Rada, 2002), caracterizándose por ser sencillo y eficiente. Por los motivos anteriormente mencionados, en este estudio se ha escogido el algoritmo k-medias.

Algoritmo k-medias: Se caracteriza por la distancia empleada en el análisis clúster denominada como euclídea, la cual se caracteriza por ser la distancia más corta entre objetos. Este algoritmo está “basado en la minimización de la distancia interna (la suma de las distancias de los patrones asignados a un agrupamiento al centroide de dicho agrupamiento)”(Berzal, 2005); por lo que, se define K centroides en cada uno de los grupos

para posteriormente seleccionar cada punto del dataset y ubicarlo en la clase de centroide más próximo.

Xu, R., y Wunsch (2009) proponen los siguientes pasos a seguir para la ejecución del algoritmo de k-medias:

- 1) Inicializar cada partición K de forma randómica o en base a centroides preexistentes. Calcular la matriz prototipo del clúster $M = [m_1 , \dots , m_K]$
- 2) Asignar cada sujeto al clúster más cercano
- 3) Recalcular la matriz prototipo del clúster en base a la partición actual
- 4) Repetir los pasos 2 y 3 hasta que no haya cambios en el centroide de los clúster (p. 144).

Las variables de entrada escogidas para la ejecución del algoritmo de k-medias referente a las actividades académicas son las variables más representativas para la clasificación como: descarga de recursos educativos, consultas a los profesores por mes, consultas a los compañeros por mes, videos vistos en youtube, y búsqueda de información académica; estas variables se encuentran agrupados en un perfil denominado: Perfil del uso del Internet en actividades académicas, en donde, se experimentó la clasificación para 2, 3, y 4 grupos, escogiéndose el grupo 3 denominado “cluster3_acad”; este grupo es escogido en base al principio que menciona Díaz De Rada (2002), que indica que los grupos que son seleccionados deben ser de fácil interpretación. En base a sus resultados, este clúster cuenta con las siguientes categorías: medio (cluster1), interactivo (clúster 2) y pasivo (clúster 3), cada uno con sus características comunes (p. 92,93).

Por otro lado, las variables que se encuentran agrupados en la clasificación denominada perfil del uso del Internet para entretenimiento son: horas que se chatea por diversión, uso de las redes sociales y videos para entretenimiento visto en youtube. Así mismo, se realizó la clasificación para 2, 3 y 4 grupos, seleccionando el grupo 3. En este perfil cada uno de los conglomerados toma el nombre de: intermedio, dinamico e inactivo (p. 95).

Discriminación.

Para verificar o evaluar los resultados que genera el algoritmo k-medias en este estudio, es necesario utilizar el análisis discriminante (discriminación) con el objetivo de analizar las diferencias significativas que existen entre los grupos respecto a un conjunto de variables medidas sobre las mismas (De la Fuente, 2011b); tomándose en consideración las recomendaciones propuestas por Díaz De Rada (1998) en la cual, se utiliza como variable dependiente al número de conglomerado resultante en la clusterización y el porcentaje de exactitud que se consigue de las asignaciones correctas.

Al aplicar el análisis discriminante ayudará analizar las diferencias significativas entre los grupos clasificados en el análisis clúster, por lo que resultó que el clúster académico escogido “cluster3_acad” (perfil para actividades académicas) tiene un porcentaje de exactitud del 94,8%, tal como indica la siguiente Tabla.

Tabla 5. Discriminación para actividades académicas de 3 grupos

		Número inicial de casos	Grupo de pertenencia pronosticado			Total
			1	2	3	
Original	Recuento	1	80	2	0	82
		2	1	111	14	126
		3	0	9	287	296
	%	1	97,6	2,4	,0	100,0
		2	,8	88,1	11,1	100,0
		3	,0	3,0	97,0	100,0

a. Clasificados correctamente el 94.8% de los casos agrupados originales.

Fuente: Elaboración propia

En cuanto al clúster de entretenimiento “cluster3_entret” (perfil para entretenimiento) escogido dio como resultado el 98% (Tabla 6) de exactitud de la clasificación; obteniendo una diferencia mínima con respecto a los demás grupos, Por tal razón, este análisis permite escoger el grupo que contenga el mayor porcentaje de exactitud como también la fácil interpretación de sus datos plasmados en los histogramas realizados.

Tabla 6. Discriminación para entretenimiento de 3 grupos

			Grupo de pertenencia pronosticado			Total
			1	2	3	
Original	Recuento	1	166	1	4	171
		2	1	85	0	86
		3	4	0	243	247
	%	1	97,1	,6	2,3	100,0
		2	1,2	98,8	,0	100,0
		3	1,6	,0	98,4	100,0

a. Clasificados correctamente el 98,0% de los casos agrupados originales.

Fuente: Elaboración propia

3.2.4. Fase de evaluación e interpretación de los datos.

Esta fase se encarga de identificar los patrones obtenidos e irrelevantes de la fase anterior, fundamentándose en algunas medidas y realizando la evaluación respectiva de los resultados obtenidos; esta fase es muy importante para la verificación de las hipótesis planteadas en la investigación, dando a conocer si estas se cumplen o no. Para ello, es necesario emplear otra técnica de minería de datos que ayuden a dar contestación a las mismas. Entre las técnicas de minería de datos que ayudan a la evaluación de los datos se encuentra la técnica de Regresión Logística; siendo esta, una técnica supervisada o predictiva que tiene como principal objetivo explicar y predecir una probabilidad de que ocurra un evento, en una variable categórica de grupos en función de diversas variables independientes o explicatorias (Taucher, 1997).

La regresión logística se divide en dos grupos: el primer grupo corresponde a la regresión logística binomial y el segundo grupo pertenece a la regresión logística multinomial; las mismas que se diferencian en el número de categorías que obtenga la variable dependiente; es decir, será binomial cuando el atributo dependiente posee dos valores categóricos (dicotómica) y será multinomial cuando el atributo dependiente esté compuesto con más de dos categorías (politómica) (Pando Fernández y San Martín Fernández, 2004)

3.2.4.1. Regresión multinomial.

En esta investigación se empleó la regresión logística multinomial para dar contestación a la hipótesis 1 (H1) y la hipótesis 2 (H2); debido a que la variable dependiente de cada una de las

hipótesis, constan de tres valores categóricos, resultados de la clusterización, tanto para perfil para actividades académicas y el perfil para entretenimiento. Además, la variable independiente (ingresos) de las dos hipótesis mencionadas consta de 5 niveles por lo que es una variable categórica nominal.

La regresión logística multinomial cuenta con varios test estadísticos que ayudan a la comprobación y ajuste del modelo seleccionado. Por lo que se realizó la comprobación de las relaciones significativas entre dos variables categóricas utilizando la técnica del Chi-cuadrado (Tabla 7) para dar contestación si se cumplen o no (en forma general) las hipótesis planteadas, en este caso se verificaron la relación significativa entre la variable independiente y la variable dependiente.

Tabla 7. Valores del Chi-cuadrado de la H1 y H2

Hipótesis	Valor del Chi-cuadrado	Grados de libertad (gl)	Nivel de significancia (Sig.)
Hipótesis 1	23.476	8	0.003
Hipótesis 2	11.285	8	0.186

Fuente: Elaboración propia

Según Torres-Díaz (2012) los coeficientes que se debe tomar en cuenta para que el modelo se ajuste a los datos en la regresión logística multinomial son:

- Chi-cuadrado del logaritmo de la verosimilitud: Determina el resultado de la resta de los valores del modelo de intercepción con el modelo ajustado, dando como resultado un modelo significativo ($p < 0,05\%$) con al menos una variable con coeficiente distinto de cero.
- Valor de Pearson y Deviance: En esta bondad de ajuste los valores obtenidos deben ser no significativos ($p > 0,05\%$) para el modelo se ajuste a los datos
- R² Nagelkerke: Explica la varianza del modelo el cual debe ser mayor al 70%.
- Prueba de Wald: Permite la comprobación de los coeficientes de las variables con un nivel de significancia $p < 0,05\%$, confirma que el coeficiente de la variable independiente es distinto de cero.
- Test de Odds Ratio: Este test permite determinar la probabilidad de que la variable dependiente pertenezca a un grupo determinado por la variable independiente, en donde el nivel de significancia de las variables debe ser menor a 0,05% (p. 96, 97, 98).

3.2.4.2. Regresión binomial.

Seguidamente se utiliza la regresión logística binomial para verificar la comprobación de la hipótesis 3 (H3) y de la hipótesis 4 (H4) relacionadas con la pregunta número dos del proyecto de investigación, estas tienen como variable dependiente (rendimiento_academico) una variable dicotómica, es decir, que toma los valores de “0” (aprobó) y “1” (no-aprobó); mientras que la variable independiente tanto para H3 y H4 consta de tres niveles cada una, la cual da como resultado una variable categórica. La Tabla 8 muestra los resultados de la prueba del Chi-cuadrado que fueron sometidas las variables de las hipótesis H3 y H4 para observar si existía una relación entre las variables independientes con las variables dependientes.

Tabla 8. Valores del Chi-cuadrado de la H3 y H4

Hipótesis	Valor del Chi-cuadrado	Grados de libertad (gl)	Nivel de significancia (Sig)
Hipótesis 3	5,660	2	0.059
Hipótesis 4	0,674	2	0,714

Fuente: Elaboración propia

En este tipo de regresión según los autores Aguayo (2007, p. 8) y Bernal (2013, p. 56) se utilizan los siguientes tests estadísticos para la comprobación del modelo que mejor se adapte a los datos, verificando si se cumplen o no las hipótesis antes señaladas, los resultados de las mismas se presentan en las páginas 99 a la 100.

- Pruebas de Omnibus: muestra un valor Chi-cuadrado que permite evaluar la hipótesis nula que indica que los coeficientes (β) de todas las variables con excepción de la constante que se encuentran contenidos en el modelo son cero; por ello, el nivel de significancia debe ser menor al 0.05% indicando que el modelo con la nueva variable introducida (variable independiente) mejora el ajuste de forma significativa con respecto a la variable dependiente.
- Prueba de Hosmer y Lemeshow: Evalúa la bondad de ajuste del modelo global, este valor no debe ser significativo para que exista un buen ajuste del modelo
- R² Nagelkerke: Verifica el porcentaje de exactitud del modelo, explicando la varianza de la variable dependiente estimada por las variables independientes. Cuanto más alto es el valor del R² más explicativo es el modelo.

- Prueba de Wald: Permite comprobar si los coeficientes de las variables independientes son significativos o no, lo que ayuda al rechazo de la hipótesis nula que indica que los coeficientes son cero.
- El test estadístico de Odds Ratio: Este test explica cuanto es más probable que suceda un evento con relación a que no acontezca ningún suceso; aquí se observa a las variables con sus respectivos intervalos de confianza.

CAPÍTULO 4: RESULTADOS

4.1. Fase de Minería de datos.

4.1.1. Datos generales de la población universitaria.

La mayor proporción de estudiantes encuestados son hombres y en menor proporción corresponde a las estudiantes mujeres. Figura 8.

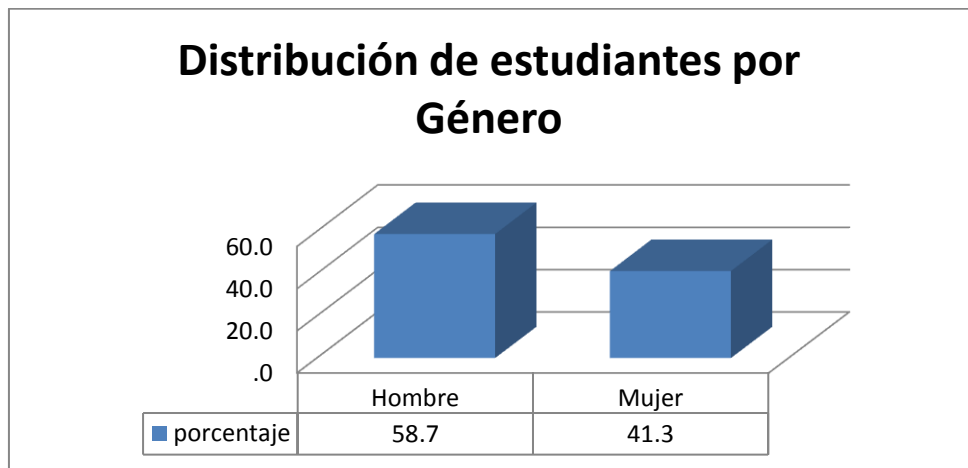


Figura 8. Distribución de estudiantes por género.

Fuente: Datos obtenidos en la investigación
Elaboración propia

El 84,5% de los estudiantes encuestados se encuentran comprendidos entre las edades 18 a 23 años de edad. Figura 9.

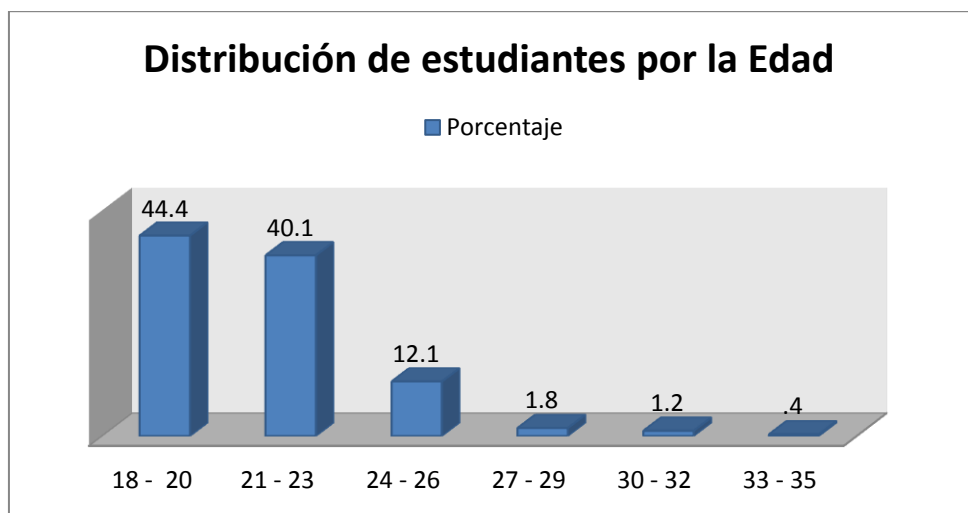


Figura 9. Distribución de estudiantes por edad.

Fuente: Datos obtenidos en la investigación
Elaboración propia

En la Figura 10 se puede observar que el mayor porcentaje de los estudiantes tienen ingresos familiares hasta 350 dólares mensuales, mientras quienes perciben ingresos familiares de más de 1500 dólares, son un porcentaje mínimo de estudiantes.

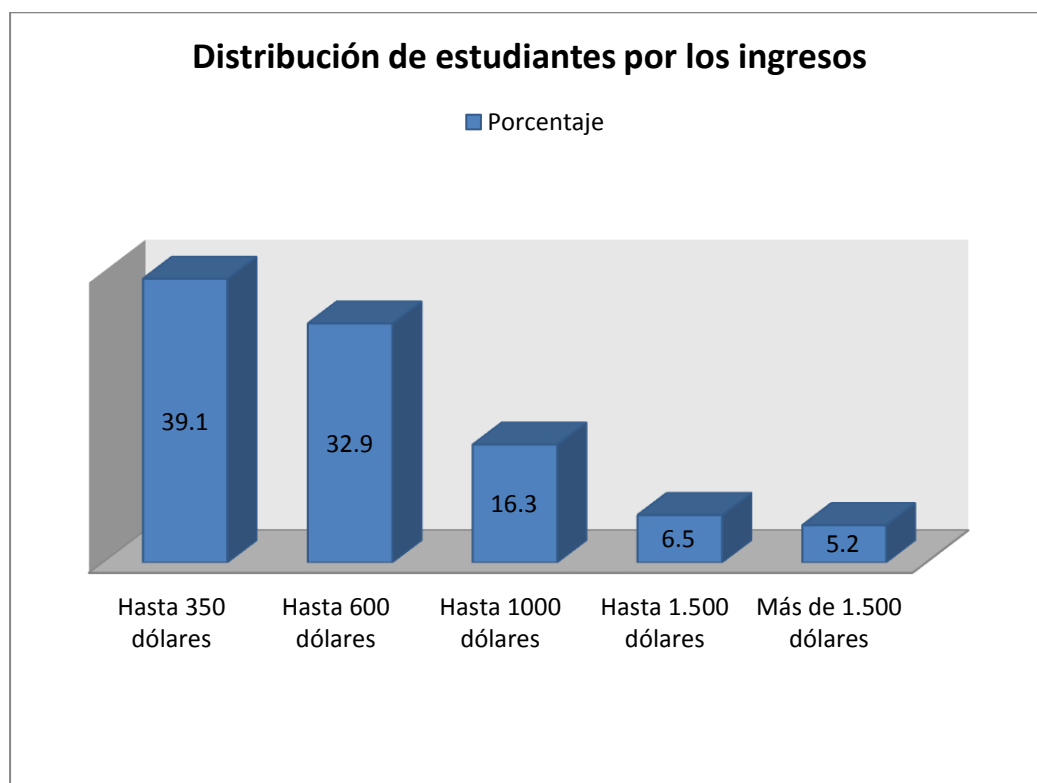


Figura 10. Distribución de estudiantes por los Ingresos familiares
Fuente: Datos obtenidos en la investigación
Elaboración propia

4.1.2. Uso de Internet en las preferencias del estudiante.

Se puede observar en la Figura 11, que el lugar habitual de conexión es desde la casa (78,4%), pero no involucra que no puede hacer uso desde otros lugares, sin embargo, el tipo de conexión en menor prioridad es la conexión desde las redes móviles (movistar, claro, cnt). El uso del Internet desde el lugar del trabajo es escaso (0,4%).

Se ha encontrado que el lugar de conexión (lug_con) esta asociado con el nivel de ingresos (ing). Por tanto, existe una relación directa y significativa entre las lug_con” e “ing”, ($\chi^2= 39.015$; $p= 0.001 < 0.05$), descartándose la hipótesis nula que indica que el nivel de ingresos no influyen en el lugar que se conecta habitualmente el estudiante a Internet.

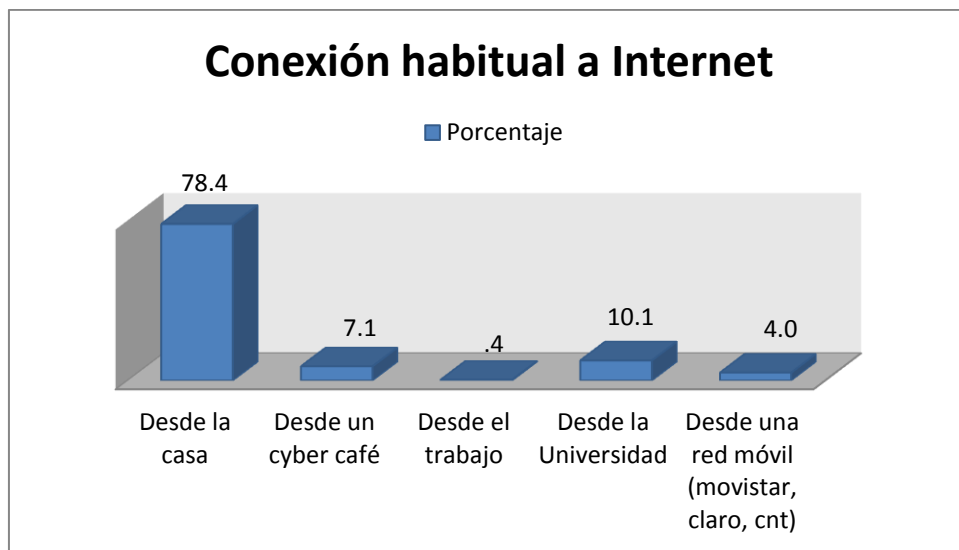


Figura 11. Conexión habitual a Internet
Fuente: Datos obtenidos en la investigación
 Elaboración propia

La mayor proporción de estudiantes se conecta los 7 días a la semana alcanzando el 55,6%, mientras que la menor proporción ingresa 1 día (1%). Tabla 9. La relación entre los niveles de ingresos con los días de conexión a la semana presentan una asociación significativa ($R_{pearson} = 0.232$; $C_{determinación} = 5.3\%$; $p = 0.000 < 0,05$), concluyendo que los niveles de ingreso influyen en los días de conexión a Internet a la semana.

Tabla 9. Días de conexión a Internet

Número de días		Frecuencia	Porcentaje	Porcentaje acumulado
Válidos	1	5	1,0	1,0
	2	12	2,4	3,4
	3	29	5,8	9,1
	4	37	7,3	16,5
	5	88	17,5	33,9
	6	53	10,5	44,4
	7	280	55,6	100,0
	Total	504	100,0	

Fuente: Datos obtenidos en la investigación
 Elaboración propia

Los niveles de conocimiento en el manejo del Internet se muestran en la Tabla 10, en ella se indica el nivel del conocimiento que tienen un estudiante en una escala de 1 a 10, se puede

observar que el porcentaje de estudiantes con un mayor nivel de conocimiento se encuentran alrededor del nivel 8, alcanzando el 28,6%; por tal razón, aproximadamente el 95% de estudiantes tienen conocimientos entre los niveles 5 y 10, mientras que los que se encuentran entre los niveles 1 y 4 no superar el 5%.

El género con el nivel de conocimiento en el manejo del Internet, tienen una relación significativa (tau-c = - 0.150; p=0.001<0.05); así mismo, existe una asociación significativa entre la variables ingresos y el nivel de conocimiento (R.pearson = 0.139; C.determinación= 1.9%; p=0.002<0.05).

Tabla 10. Nivel de conocimiento en uso del Internet

		Frecuencia	Porcentaje	Porcentaje acumulado
Válidos	1	2	,4	,4
	2	2	,4	,8
	3	9	1,8	2,6
	4	12	2,4	5,0
	5	61	12,1	17,1
	6	51	10,1	27,2
	7	111	22,0	49,2
	8	144	28,6	77,8
	9	68	13,5	91,3
	10	44	8,7	100,0
	Total	504	100,0	

Fuente: Datos obtenidos en la investigación
Elaboración propia

El tiempo de horas de conexión diaria para sus actividades académicas o de entretenimiento, se ve reflejada en la Tabla 11, donde el mayor porcentaje de estudiantes se conecta de 1 a 3 horas al día (48%), y en menor proporción se conectan de 13 a 15 horas diarias (0,2%).

Existe una asociación significativa entre la edad y las horas que se conecta al Internet el estudiante (R.pearson= 0.156 ; C.determinación= 2.4%; p=0.000<0.05), como también, se presenta un nivel de significancia entre los ingresos y las horas que se conecta al Internet el estudiante en el día (tau-c= 0.075; p=0.030<0.05).

Tabla 11. Horas que se conecta al día

		Frecuencia	Porcentaje	Porcentaje acumulado
Válidos	1,00 - 3,00	242	48,0	48,0
	4,00 - 6,00	177	35,1	83,1
	7,00 - 9,00	49	9,7	92,9
	10,00 - 12,00	29	5,8	98,6
	13,00 - 15,00	7	1,4	100,0
	Total	504	100,0	

Fuente: Datos obtenidos en la investigación
Elaboración propia

Al realizar el levantamiento de información desde hace cuantos años se conecta el estudiante a Internet, la mayor proporción se concentra en los años de 4 a 6 (49%), mientras que la menor proporción están alrededor de los 13 a 15 años de experiencia en Internet. Tabla 12.

Los resultados en la investigación presentan una relación entre la edad (eda), los ingresos(ing) con los años de experiencia (año_exp) en el uso del Internet. La primera relación entre las variables “eda” y “año_exp” tienen un valor (R.pearson = 0.123 ; C.determinacion= 1.51%) con un nivel de significancia $p < 0,05$. La segunda relación entre las variables “ing” y “año_exp” ($\tau\text{-}c = 0.068$; $p = 0.044 < 0.05$) tienen un nivel de significancia $p < 0,005$; por tanto, existe una relación significativa en ambos casos.

Tabla 12. Años de experiencia en Internet

		Frecuencia	Porcentaje	Porcentaje acumulado
Válidos	1,00 - 3,00	111	22,0	22,0
	4,00 - 6,00	247	49,0	71,0
	7,00 - 9,00	99	19,6	90,7
	10,00 - 12,00	45	8,9	99,6
	13,00 - 15,00	2	,4	100,0
	Total	504	100,0	

Fuente: Datos obtenidos en la investigación
Elaboración propia

Con respecto a los estudiantes que tienen blog, el 16,5% respondieron afirmativamente, mientras que los estudiantes que tienen cuenta en youtube tienen un porcentaje del 56,7%, por otro lado, los estudiantes que poseen una cuenta en del.icio.us contienen un porcentaje bajo del 4,6%. Figura 12.

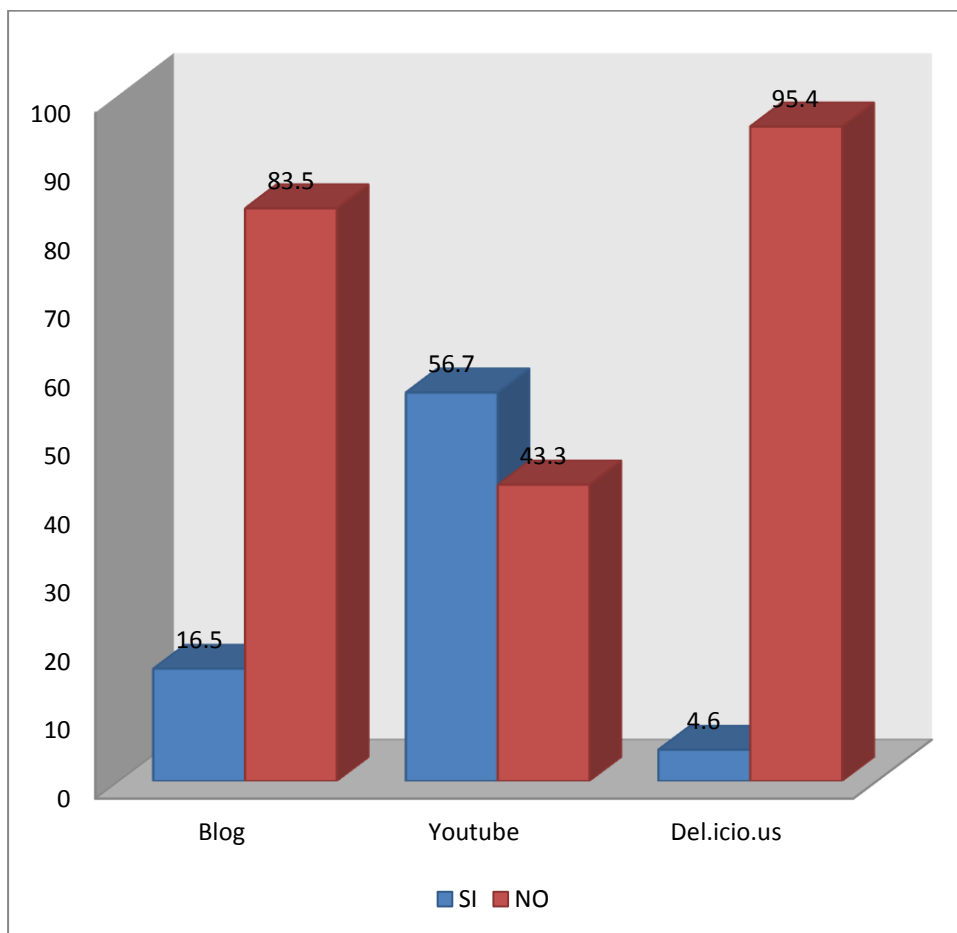


Figura 12. Estudiantes que tienen blog, cuenta en youtube y en del.icio.us.

Fuente: Datos obtenidos en la investigación

Elaboración propia

La Tabla 13 muestra el nivel de uso de los dispositivos que manejan los estudiantes, en donde se puede observar que la mayor parte de los estudiantes tienen un nivel de “UNO” (valor más bajo) en el uso de los dispositivos como el smartphone (38,5%), teléfono móvil con acceso a Internet (24,8%), tablet (51,6%), el teléfono móvil sin acceso a Internet (38,5%), cámara digital (37,9%) y el iPod (52,6%); se consigue evidenciar que en el uso del computador portátil existe un alto índice de estudiantes (38,3%) con un nivel de 10 (valor más alto).

Tabla 13. Nivel de uso de los dispositivos móviles

nivel de uso	Smartphone	Teléfono móvil con acceso a Internet	Tablet (iPod; Galaxy Tab, Kindle, etc.)	Computador portátil	Teléfono móvil sin acceso a Internet	Cámara digital	iPod / MP3 Player
	Porcentaje						
1	38,5	24,8	51,6	10,1	38,5	37,9	52,6
2	2,8	2,8	2,8	2,0	6,0	6,5	6,7
3	5,4	4,0	4,2	2,8	6,2	6,7	3,4
4	4,0	3,6	4,2	4,2	3,0	5,4	4,2
5	7,3	7,9	6,2	8,3	7,5	7,5	6,0
6	4,2	4,2	2,0	4,6	2,0	4,8	3,0
7	6,2	5,8	5,0	5,6	3,6	6,0	3,8
8	8,3	10,3	5,6	12,3	5,4	5,4	5,2
9	8,1	12,1	7,1	11,9	7,5	6,0	4,4
10	15,3	24,6	11,5	38,3	20,4	13,9	10,9

Fuente: Datos obtenidos en la investigación
Elaboración propia

La utilización de los dispositivos móviles como el smartphone, teléfono móvil con acceso a Internet, teléfono móvil sin acceso a Internet, la tablet y la cámara digital, están directamente relacionadas con la edad, dándonos como resultado una asociación significativa ($p < 0.005$).

De la misma forma el uso de los dispositivos como el smartphone, telefono movil con acceso a Internet, el uso del computador portatil, la tablet, cámara digital y el uso del Ipod tienen relación con el nivel de ingresos de la familia ($p = 0.000 < 0.05$), concluyendo una dependencia significativa en donde el nivel de ingresos que tiene la familia del estudiante influyen directamente en el uso de los dispositivos móviles.

4.1.3. Uso del Internet en las actividades académicas.

La Figura 13, muestra que el 81,8% de los estudiantes ingresan a la plataforma virtual de 1 a 3 veces a la semana.

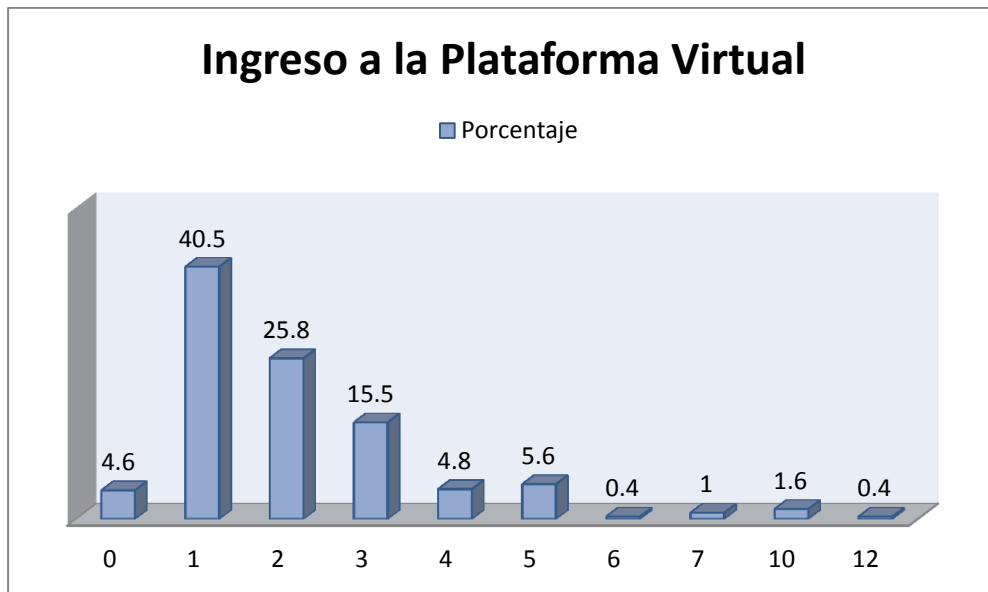


Figura 13. Ingreso a la plataforma virtual
Fuente: Datos obtenidos en la investigación
 Elaboración propia

La Tabla 14 indica que el 52,4% de los estudiantes realizan entre 1 y 5 consultas a su profesores al mes, además, existe un considerable porcentaje (14,7 %) de estudiantes que no realizan consultas a su profesor.

El género afecta directa y significativamente a la variable dependiente consultas que se realiza al profesor ($\tau\text{-}c=0.117$) con un nivel de significancia ($p=0.019<0.05$).

Tabla 14. Consultas realizadas a los profesores por mes

		Frecuencia	Porcentaje	Porcentaje acumulado
Válidos	0,00	74	14,7	14,7
	1,00 - 5,00	264	52,4	67,1
	6,00 - 10,00	95	18,8	85,9
	11,00 - 15,00	29	5,8	91,7
	16,00 - 20,00	27	5,4	97,0
	21,00 - 25,00	6	1,2	98,2
	26,00 - 30,00	9	1,8	100,0
	Total	504	100,0	

Fuente: Datos obtenidos en la investigación
 Elaboración propia

El 43,7% de los estudiantes realizan de una a cinco consultas a los compañeros sobre temas académicos por mes. Tabla 15.

Se halla una asociación significativa entre la edad y las consultas que se realiza a los compañeros (R.pearson= -0.096 ; C.determinacion= 0.92% ; p= 0.031 < 0.05).

Tabla 15. Consultas realizadas a los compañeros por mes

		Frecuencia	Porcentaje	Porcentaje acumulado
Válidos	0,00	47	9,3	9,3
	1,00 - 5,00	220	43,7	53,0
	6,00 - 10,00	123	24,4	77,4
	11,00 - 15,00	39	7,7	85,1
	16,00 - 20,00	47	9,3	94,4
	21,00 - 25,00	7	1,4	95,8
	26,00 - 30,00	21	4,2	100,0
	Total	504	100,0	

Fuente: Datos obtenidos en la investigación
Elaboración propia

Las descargas mensuales de los recursos educativos desde la plataforma virtual de acuerdo a la investigación se encuentran entre 1 a 3 descargas al mes (34,5%) seguidamente el 26,2% de los estudiantes no realizan ninguna descarga de recursos educativos. Tabla 16.

Tabla 16. Descargas de recursos educativos por mes

		Frecuencia	Porcentaje	Porcentaje acumulado
Válidos	0,00	132	26,2	26,2
	1,00 - 3,00	174	34,5	60,7
	4,00 - 6,00	81	16,1	76,8
	7,00 - 9,00	31	6,2	82,9
	10,00 - 12,00	56	11,1	94,0
	13,00 - 15,00	15	3,0	97,0
	16,00 - 18,00	1	,2	97,2
	19,00 - 21,00	14	2,8	100,0
	Total	504	100,0	

Fuente: Datos obtenidos en la investigación
Elaboración propia

La Tabla 17 indica que alrededor del 52,1% de los estudiantes observan de 1 a 6 videos académicos a través del youtube cada mes.

La variable independiente “género” esta asociada significativamente ($p=0.000<0.05$) con la variable dependiente “videos vistos en youtube” ($\tau\text{-}c= -0.190$).

Tabla 17. Videos académicos vistos en youtube cada mes

		Frecuencia	Porcentaje	Porcentaje acumulado
Válidos	0,00	43	8,5	8,5
	1,00 - 3,00	105	20,4	29,4
	4,00 - 6,00	160	31,7	61,1
	7,00 - 9,00	22	4,4	65,5
	10,00 - 12,00	97	19,2	84,7
	13,00 - 15,00	35	6,9	91,7
	16,00 - 18,00	3	,6	92,3
	19,00 - 21,00	39	7,7	100,0
	Total	504	100,0	

Fuente: Datos obtenidos en la investigación
Elaboración propia

A continuación, en la Figura 14 se muestra que la mayor parte de los estudiantes en este estudio no participan de los foros virtuales mensuales.

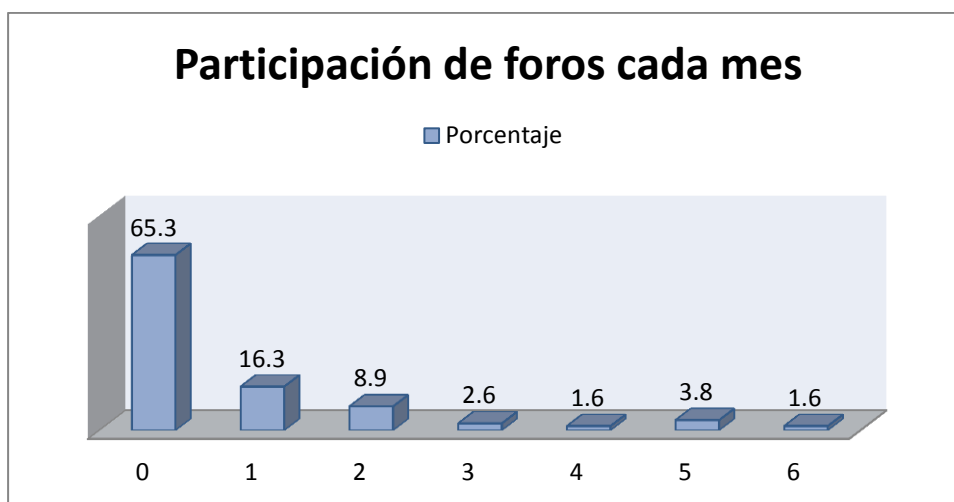


Figura 14. Participación de foros cada mes.

Fuente: Datos obtenidos en la investigación
Elaboración propia

Se puede observar en la Figura 15, que la mayor proporción de los estudiantes no postean ni realizan tweets académicos en las redes sociales cada mes en el ámbito académico, seguidamente el 30,1% de estudiantes realizan de 1 a 4 posts o tweets al mes.

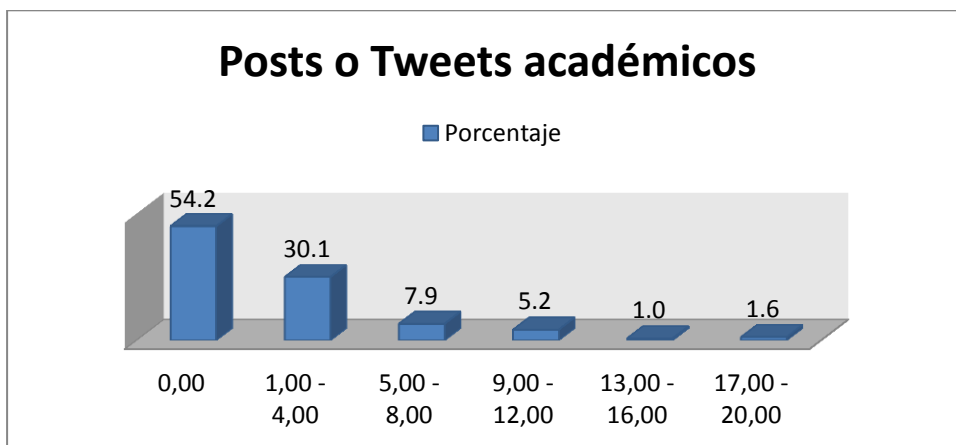


Figura 15. Participación de post o tweets cada mes.

Fuente: Datos obtenidos en la investigación

Elaboración propia

Se encontró que la mayor parte de los estudiantes chatean por temas académicos de 1 a 4 horas mensuales. Figura 16.

Existe una relación significativa entre la edad y las horas que se emplean para el chat destinado a temas académicos ($R_{\text{pearson}} = -0.136$; $C_{\text{determinación}} = 1.84\%$; $p = 0.002 < 0.05$).

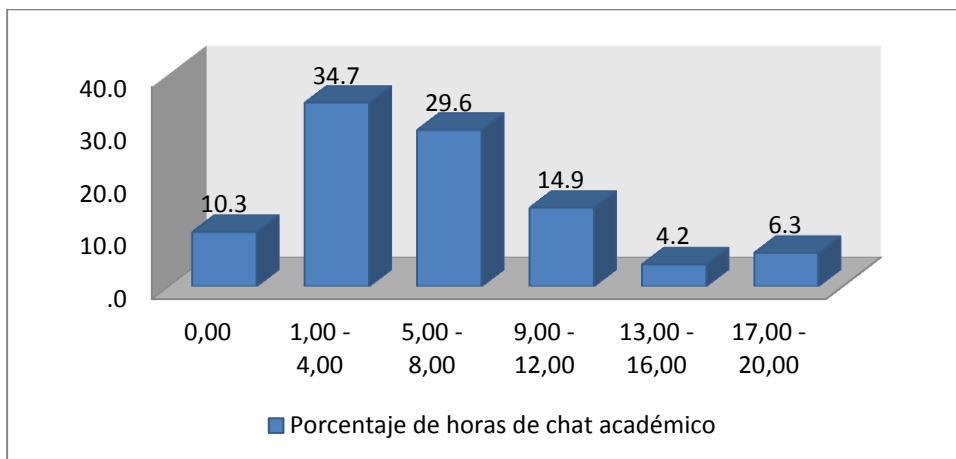


Figura 16. Chat destinado a temas académicos por mes.

Fuente: Datos obtenidos en la investigación

Elaboración propia

El 43,2% de los estudiantes buscan información académica en Internet de 1 a 10 horas mensuales. Ver la Tabla 18.

Tabla 18. Búsqueda de información académica en Internet al mes

		Frecuencia	Porcentaje	Porcentaje acumulado
Válidos	0,00	11,0	2,2	2,2
	1,00 - 10,00	218	43,2	45,4
	11,00 - 20,00	151	30,0	75,4
	21,00 - 30,00	72	14,3	89,7
	31,00 - 40,00	20	4,0	93,7
	41,00 - 50,00	19	3,8	97,4
	51,00 - 60,00	13	2,6	100,0
	Total	504	100,0	

Fuente: Datos obtenidos en la investigación
Elaboración propia

La mayor parte de los estudiantes ingresan a la biblioteca virtual entre 1 y 5 horas mensuales, además existe un porcentaje considerable (28%) de estudiantes que no ingresan a la biblioteca virtual. Tabla 19.

Tabla 19. Ingreso a la biblioteca virtual

		Frecuencia	Porcentaje	Porcentaje acumulado
Válidos	0,00	141	28,0	28,0
	1,00 - 5,00	250	49,6	77,6
	6,00 - 10,00	71	14,1	91,7
	11,00 - 15,00	21	4,2	95,8
	16,00 - 20,00	11	2,2	98,0
	21,00 - 25,00	1	,2	98,2
	26,00 - 30,00	9	1,8	100,0
	Total	504	100,0	

Fuente: Datos obtenidos en la investigación
Elaboración propia

4.1.4. Uso del Internet para entretenimiento.

En este apartado, la Tabla 20 muestra que la mayor concentración de estudiantes (46,6%) utiliza de 1 a 5 horas de chat semanal para entretenimiento o diversión.

Se ha encontrado que la variable edad con las horas que se chatea por diversión, existe una relación significativa (R.pearson= -0.188; C.determinación= 3.53%; p=0.000<0.05).

Tabla 20. Horas de chat semanal por diversión

		Frecuencia	Porcentaje	Porcentaje acumulado
Válidos	0,00	6	1,2	1,2
	1,00 - 5,00	235	46,6	47,8
	6,00 - 10,00	157	31,2	79,0
	11,00 - 15,00	37	7,3	86,3
	16,00 - 20,00	24	4,8	91,1
	21,00 - 25,00	25	5,0	96,0
	26,00 - 30,00	20	4,0	100,0
	Total	504	100,0	

Fuente: Datos obtenidos en la investigación
Elaboración propia

La Tabla 21 muestra que el 36,3% de los estudiantes en nuestra investigación emplean de 1 a 5 horas semanales en redes sociales.

La edad y el género presenta una relación significativa con el uso de las redes sociales (p<0.05%) cuyos valores son: (R.pearson = -0.184; C.determinación= 3.38%) y (tau-c= -0.116) correspondientemente.

Tabla 21. Horas semanales empleadas en redes sociales

		Frecuencia	Porcentaje	Porcentaje acumulado
Válidos	0,00	7	1,4	1,4
	1,00 - 5,00	183	36,3	37,7
	6,00 - 10,00	164	32,5	70,2
	11,00 - 15,00	46	9,1	79,4
	16,00 - 20,00	39	7,7	87,1
	21,00 - 25,00	24	4,8	91,9
	26,00 - 30,00	27	5,4	97,2
	31,00 - 35,00	5	1,0	98,2
	36,00 - 40,00	9	1,8	100,0
	Total	504	100,0	

Fuente: Datos obtenidos en la investigación
Elaboración propia

En la investigación alrededor del 59,3% de los estudiantes no hacen uso de sus horas para los juegos en línea cada semana (Tabla 22). Además, se encontró una asociación significativa ($p < 0,05\%$) entre el género y los juegos en línea ($\text{tau-c} = - 0.144$).

Tabla 22. Horas semanales dedicadas a los juegos en línea

		Frecuencia	Porcentaje	Porcentaje acumulado
Válidos	0,00	299	59,3	59,3
	1,00 - 5,00	142	28,2	87,5
	6,00 - 10,00	34	6,7	94,2
	11,00 - 15,00	4	,8	95,0
	16,00 - 20,00	10	2,0	97,0
	21,00 - 25,00	7	1,4	98,4
	26,00 - 30,00	8	1,6	100,0
	Total	504	100,0	

Fuente: Datos obtenidos en la investigación
Elaboración propia

En la Figura 17, se puede observar que la mayor concentración de estudiantes descargan de 1 a 3 horas semanales: música, videos y programas. Se descubrió que entre la variable “eda” y la variable “e_des_con”, existe una relación significativa ($p < 0.005$), por lo que se deduce que la edad influye en las horas que emplea el estudiante para descargar música, videos y programas de entretenimiento.

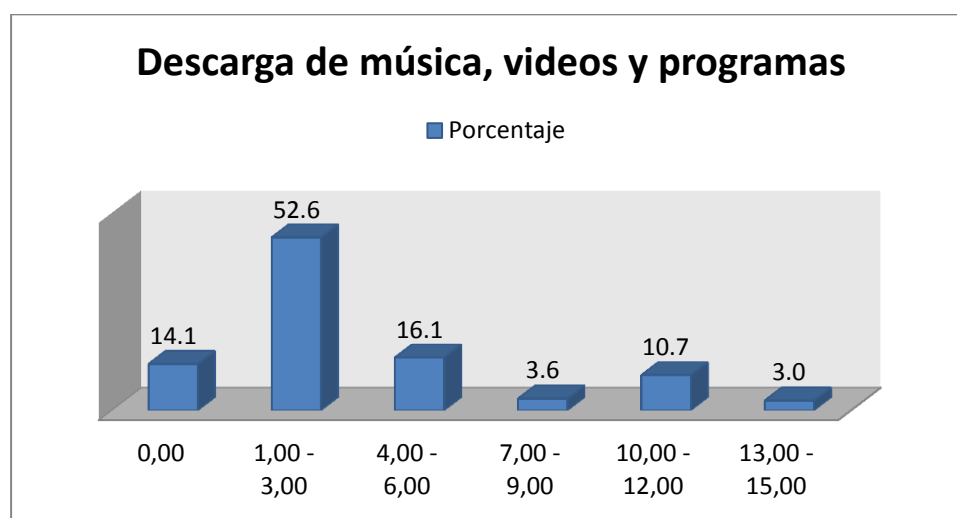


Figura 17. Horas de descargas de música, videos y programas.

Fuente: Datos obtenidos en la investigación
Elaboración propia

La Tabla 23 muestra que la mayor proporción de estudiantes (53,4%) observan de 1 a 5 videos en youtube a la semana. Se ha encontrado una relación significativa entre el género y las horas que utiliza el estudiante para mirar videos en youtube ($\tau\text{-}c = -0,209$) con un nivel de significancia $p < 0,05$.

Tabla 23. Videos para entretenimiento vistos en youtube a la semana

		Frecuencia	Porcentaje	Porcentaje acumulado
Válidos	0,00	38	7,5	7,5
	1,00 - 5,00	269	53,4	60,9
	6,00 - 10,00	117	23,2	84,1
	11,00 - 15,00	28	5,6	89,7
	16,00 - 20,00	32	6,3	96,0
	21,00 - 25,00	9	1,8	97,8
	26,00 - 30,00	11	2,2	100,0
	Total	504	100,0	

Fuente: Datos obtenidos en la investigación
Elaboración propia

La mayor proporción de seguidores en las redes sociales se encuentran entre el intervalo 0 y 1375, por lo que, la red social twitter posee el 99,6%, facebook cuenta con el 92,5% y linkedIn tienen el 99,8% de seguidores; mientras que en los demás rangos es todo lo contrario, existe una escasa concentración de seguidores en las redes sociales (Figura 18).

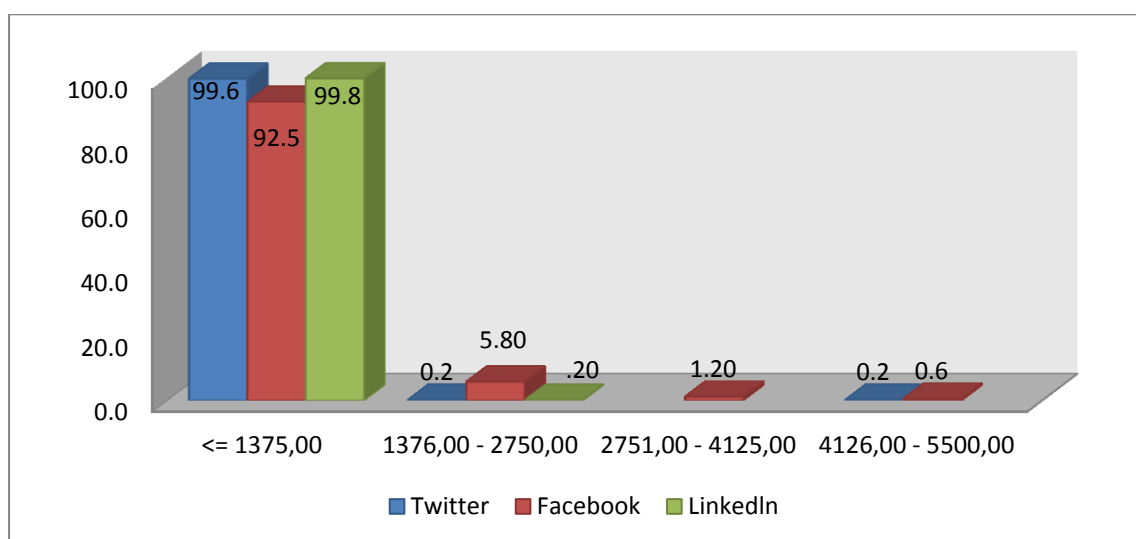


Figura 18. Seguidores en twitter, facebook y linkedIn.

Fuente: Datos obtenidos en la investigación
Elaboración propia

4.1.5. Generalidades sobre el rendimiento académico.

Para el presente estudio el rendimiento se lo midió en base a la materias que se aprueba en el módulo matriculado; esta variable (rendimiento_académico) se construyó tomando en cuenta que en la universidad investigada no se trabaja con las materias por créditos, sino que tiene la particularidad de que los estudiantes deben aprobar todas las asignaturas matriculadas, caso contrario deben repetir el módulo. Esta variable da como resultado una variable dicotómica o binaria, ya que puede tomar dos valores posibles (0 ó 1); para este caso el valor de “0” representa los estudiantes que han aprobado el módulo y el valor “1” representa los estudiantes que no han aprobado el módulo.

En este apartado se puede observar en la Figura 19, que alrededor del 96% de los estudiantes han aprobado el módulo, mientras que solo el 4% no lo hicieron.

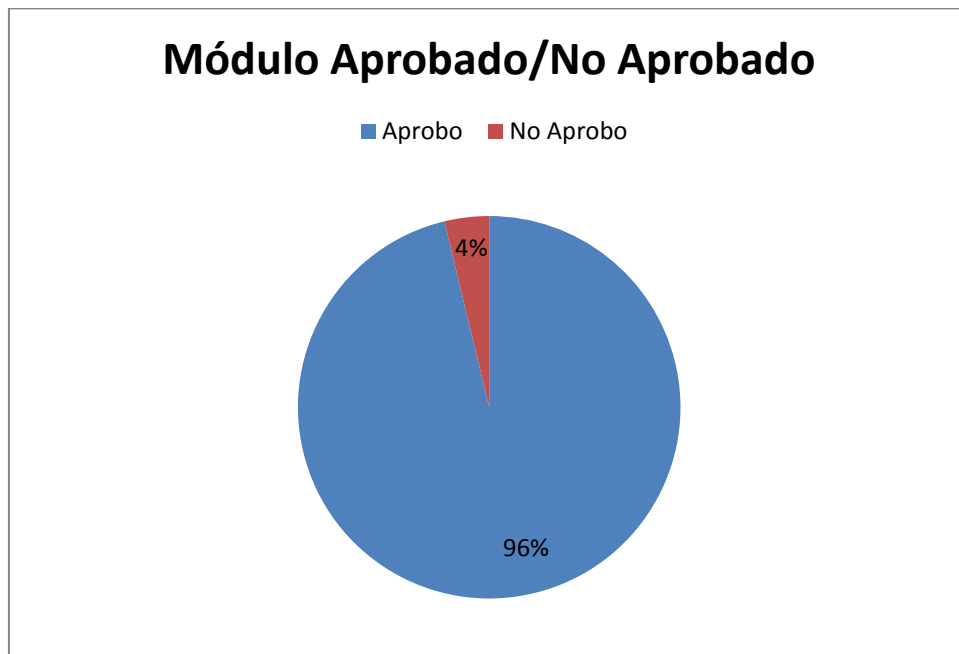


Figura 19. Módulo aprobado/no aprobado (rendimiento académico).

Fuente: Datos obtenidos en la investigación
Elaboración propia

El rendimiento académico no presenta una relación significativa con el género, la edad y el nivel de ingresos ($p > 0,05\%$), concluyendo que el rendimiento académico es independiente de la edad, el género y el nivel de ingresos que tiene el estudiante.

4.1.6. Reconocimiento de patrones.

En este apartado, se examinará la reducción de variables mediante la técnica del análisis factorial para posteriormente realizar las respectivas clasificaciones en torno a sus características comunes, como también a las diferencias que muestran los estudiantes sobre las actividades académicas, de entretenimiento y el nivel de uso de los dispositivos tecnológicos; estas clasificaciones que se encuentran agrupadas por categorías o grupos que toman el nombre de perfiles. Los perfiles en estudio son: perfil para actividades académicas, perfil para entretenimiento y perfil del nivel de uso de dispositivos.

4.1.6.1. Reducción de variables: Factorización.

Para realizar la reducción de un conjunto de variables a un número pequeño de nuevas variables se aplicó en primera instancia la técnica de minería de datos “Factorización” en todos sus perfiles, la misma que ayuda a trabajar con pocas variables, permitiendo facilitar la interpretación de los resultados. Sin embargo, de acuerdo a las condiciones explicadas en la metodología que deben cumplirse para la utilización de esta técnica, como resultado se obtuvieron que los valores no fueron factibles, ya que el valor de la determinante de la matriz de correlaciones no es cercana a cero, la medida muestral KMO dio como resultado un valor bajo, el test de esfericidad de Barlett si es significativo, pero el valor de la varianza total no llega alrededor del 70%, lo que indica que no es posible emplear esta técnica tanto en las actividades académicas como para entretenimiento (Anexo 6 y 7); concluyendo que se debe utilizar directamente el método de análisis clúster.

4.1.6.2. Perfil para actividades académicas.

Este perfil se desarrolló en base a los usos de la tecnología que se han dado para las actividades académicas de los estudiantes. El procedimiento que se utilizará para agrupar a los estudiantes de acuerdo a sus características comunes se realiza en tres etapas: Primera etapa: se elige las variables más representativas para la investigación. La segunda etapa es denominada como la clasificación y verificación, aquí se utiliza las variables que se obtuvieron como resultado de la clusterización, estas clasificaciones son verificadas con el análisis discriminante. Y finalmente la tercera etapa tenemos la interpretación de los grupos, describiéndolo los grupos resultantes.

La clasificación se desarrolló con el fin de contar con conglomerados fácil de interpretar. De las variables descritas a continuación se escogieron cuatro variables significativas para la investigación.

- ¿Cuántas veces ingresa a la plataforma virtual de sus 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?

Elección de variables.

El procedimiento para elegir las variables mas representativas sugerido para el presente estudio es el análisis de estadística de dependencia central como la mediana, ya que la misma permite identificar los valores mas distintivos de los datos. Por tanto, se tomo los valores más altos de cada una de las variables que estan inmersas en el perfil académico. De las 10 variables empleadas originalmente se redujieron a 4 variables de entrada. Las variables resultantes fueron: descarga de recursos educativos, consultas a los profesores por mes, consultas a los compañeros por mes, videos vistos en youtube, y búsqueda de información académica.

Clasificación y verificación.

Para la clasificación de los grupos de acuerdo a sus características habituales para las actividades académicas, se aplicó el algoritmo de clusterización k-medias, la misma se utiliza para muestras grandes como en la presente investigación. Este análisis se lo realizó utilizando las recomendaciones del programa SPSS (version 17). En este estudio se construyeron

clasificaciones para los grupos de 2, 3 y 4 (Anexo 7, 8 y 9).

Para la comprobación o verificación de los grupos se ha manejado el análisis discriminante, sobre las categorizaciones resultantes, utilizando como variable dependiente el número de grupo generado en el análisis clúster y como variables independientes las variables elegidas como más representativas. En la Tabla 24, se muestra el porcentaje de nivel de precisión como resultado del análisis discriminante de cada una de los grupos.

Tabla 24. Porcentaje de precisión de la clasificación (académica)

Numero de grupos	Porcentaje de Exactitud
2	94,6 %
3	94,8%
4	98,4 %

Fuente: Elaboración propia

El porcentaje del grupo 3 es menor en comparación a los porcentajes del grupo 2 y 4, sin embargo, se escogió el grupo 3 por su fácil interpretación de los datos, siendo este uno de los principios que menciona el autor Díaz De Rada (2002) en su investigación. Las variables que mayor discriminación tienen en la primera función son: la búsqueda de información, las consultas al profesor y las consultas a los compañeros por mes. En la segunda función la mayor capacidad de discriminación la poseen las variables: consultas a los compañeros, consultas al profesor y recursos educativos (Tabla 25).

Tabla 25. Nivel de discriminación para actividades académicas

Variables	Función Discriminante	
	1	2
¿Aproximadamente cuántos recursos educativos descarga de la plataforma virtual cada mes?	,016	,278
¿Aproximadamente cuántas consultas les hace a sus profesores cada mes?	,209	,496
¿Aproximadamente cuántas consultas realiza a sus compañeros por mes?	,210	,724
¿Aproximadamente cuántos videos académicos mira en youtube cada mes?	,045	,188
¿Aproximadamente cuántas horas busca información académica en Internet cada mes?	,925	-,412

Fuente: Elaboración propia

Descripción de los grupos.

En la Figura 20, se presentan los centroides de cada variable escogida para este perfil.

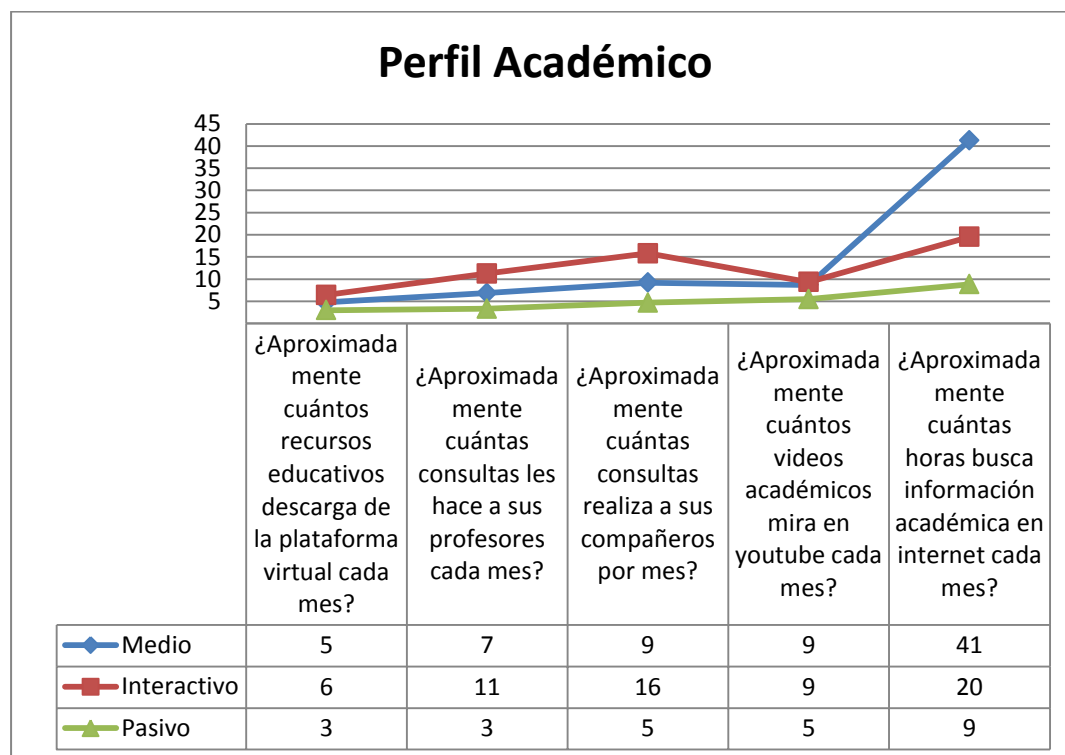


Figura 20. Perfil del uso del Internet en el aspecto académico
Fuente: Elaboración propia

El grupo 1 llamado “medio”, se conforma con el 16,3% de estudiantes y se caracteriza por tener valores medios en casi todas sus variables, excepto en la variable de búsqueda de información académica al mes, indicando que tiene un alto porcentaje; es decir, que los estudiantes que pertenecen a este grupo deben realizar al menos 5 descargas de recursos educativos, siete consultas al profesor, al menos 9 consultas a sus compañeros, observan 9 videos académicos en youtube y dedican un promedio de 41 horas en la búsqueda de información académica al mes.

El grupo 2 llamado “interactivo”, tiene alrededor del 25% de estudiantes que se caracterizan por tener un nivel más alto en: descargas de recursos educativos (6 descargas), 11 consultas a los profesores, 16 consultas a los compañeros al mes, y un nivel medio de 20 horas en la búsqueda de información académica en comparación del grupo uno y tres, pero con un promedio igual de 9 horas en mirar videos académicos en Internet que el grupo uno.

El grupo 3 llamado “pasivo”, abarca el 58,1% de estudiantes diferenciándose del grupo 1 “medio” y 2 “interactivo” por tener un nivel muy bajo en todas sus variables, es decir, los estudiantes que pertenecen a este grupo descargan solamente tres recursos educativos, realizan al menos tres consultas a sus profesores y cinco consultas a sus compañeros al mes; además, de realizar un promedio muy bajo de 9 horas en búsqueda de información académica al mes.

4.1.6.3. Perfil para entretenimiento.

Para la clasificación de los grupos en función del uso de Internet para el ámbito de entretenimiento se lo realizó tomando en cuenta el desarrollo de socialización y diversión que fomenta el estudiante al hacer uso del mismo. Por tal razón los pasos que se siguieron para su clasificación son: Elección de las variables más representativas, clasificación y verificación e interpretación del grupo que se tomó como resultante. Este perfil cuenta con las siguientes variables de donde se escogieron a las de mayor relevancia.

- ¿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?

Elección de variables.

Para elegir la cantidad de variables más importantes para esta investigación, se utilizó análisis de estadística de dependencia central como la mediana, escogiendo los valores más altos e importantes para este estudio, esta técnica permite que al manejar menos variables sea mejor la interpretación de los resultados. Este perfil cuenta con 5 variables de las cuales se eligieron a 3 variables que son: Horas que se chatea por diversión, uso de las redes sociales y videos para entretenimiento visto en youtube.

Clasificación y verificación.

Se utilizó el algoritmo k-medias perteneciente a la técnica de clusterización para clasificar a los estudiantes por grupos con similitudes usuales, siguiendo los pasos en SPSS (versión 17). El

proceso de clasificación muestra varias agrupaciones, el objetivo, al igual que en las secciones anteriores, es contar con una clasificación que sea fácil de interpretarla; para este estudio se hizo la clasificación para 2, 3, y 4 grupos. (Anexo 10, 11 y 12).

Cada una de las clasificaciones fueron sometidas a un análisis discriminante para la verificación de su porcentaje de precisión, tal como muestra la Tabla 26; se ha escogido el grupo tres apto en el aspecto interpretativo, en vez de aquella con el nivel de precisión más alto, aunque la diferencia en su nivel de precisión son mínimos.

Tabla 26. Porcentaje de precisión de la clasificación (entretenimiento)

Número de grupos	Porcentaje de exactitud
2	98,2 %
3	98,0 %
4	98,4 %

Fuente: Elaboración propia

En la función 1, las variables que tienen mayor capacidad de discriminación son: las redes sociales y las horas de chat por diversión, mientras que en la segunda función discriminante las variables que mejor discriminan son: los videos vistos en youtube y las horas de chat por diversión (Tabla 27).

Tabla 27. Nivel de discriminación para entretenimiento

Variables	Función Discriminante	
	1	2
¿Aproximadamente cuántas horas a la semana chatea por diversión?	,592	,268
¿Aproximadamente cuántas horas a la semana utiliza redes sociales?	,811	-,426
¿Aproximadamente cuántos videos para entretenimiento mira en youtube cada semana?	,352	,864

Fuente: Elaboración propia

Descripción de los grupos.

La clasificación final divide a los estudiantes en tres grupos en el ámbito de entretenimiento observar en la Figura 21.

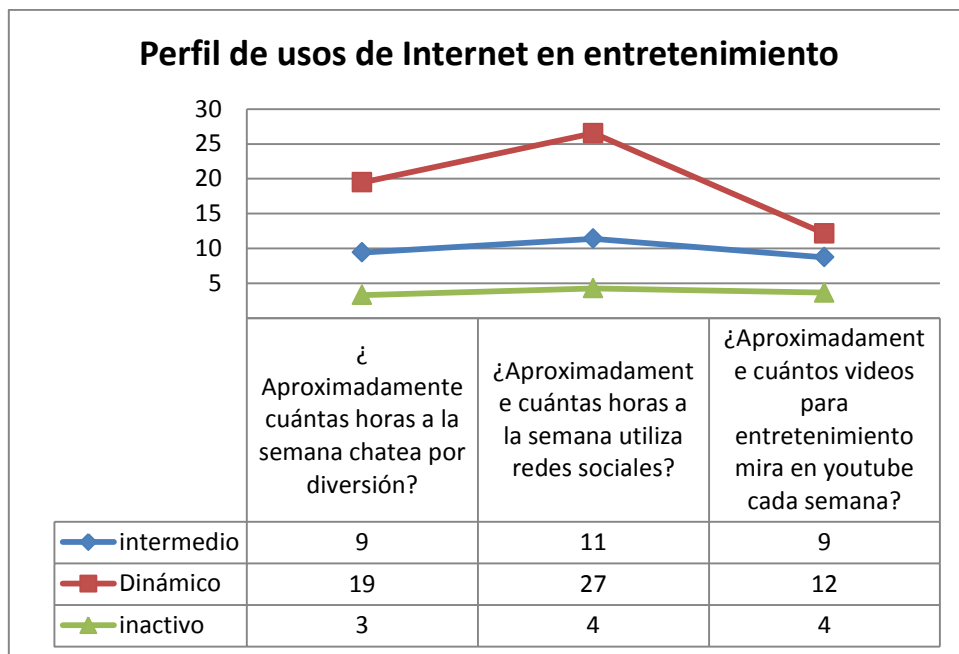


Figura 21. Perfil del uso de Internet para entretenimiento.
Fuente: Elaboración propia

El grupo 1 considerado como “intermedio” está conformado por el 33,9% de estudiantes, se caracteriza por tener valores intermedio de uso de Internet como al menos 9 horas que utiliza el chatea por diversion y alrededor de 9 videos que mira en youtube cada semana; pero haciendo énfasis en una diferencia mínima en la utilización de las redes sociales con un promedio de 11 horas semanales, en relación a los demás grupos existentes.

El grupo 2 considerado como “dinámico” se conforma por el 17,1% de estudiantes y tiene como característica primordial un valor mayor en las distintas variables que se miden como entrada, motivo por el cual este perfil de usuario tiene mas uso de horas de chat por diversion (19 horas), como también emplea 27 horas en las redes sociales y finalmente observa 12 videos en youtube, resultando un promedio mayor en comparacion que los grupos uno y tres.

El grupo 3 considerado como “inactivo” tiene alrededor del 49 % de estudiantes los que se caracterizan por tener el nivel más bajo en todas su variables con respecto a los grupos 1 y 2; considerando un tipo de usuario, que emplea tres horas de chat, cuatro horas en utilizar las redes sociales y un promedio de cuatro videos que mira en youtube por diversión; concluyendo algunas suposiciones, como por ejemplo que el entretenimiento es poco importante para los estudiantes inmersos en este grupo, no cuentan con las posibilidades tecnológicas a la mano, o

simplemente emplean su tiempo en otras actividades.

4.1.6.4. Perfil del nivel de uso de dispositivos.

Este perfil se refiere al nivel de uso de dispositivos que maneja el estudiante para comunicarse, divertirse o utilizarlo en el aprendizaje educativo, con la ventaja de hacerlo desde cualquier lugar que se encuentre el usuario. Los estudiantes que cuentan con el acceso a los dispositivos inciden significativamente en el nivel de conocimiento del estudiante, aportando experiencia, habilidades y destrezas en su manejo. Las variables que corresponden a este perfil se encuentran en una escala de 1 a 10 en cuanto a su uso, por lo que el 1 significa que el estudiante tiene un nivel escaso para el uso del dispositivo y el 10 significa que el estudiante maneja al máximo el uso de los dispositivos.

- Smartphone con cámara fotográfica y acceso a Internet
- Teléfono móvil con acceso a Internet
- Teléfono móvil sin acceso a Internet
- Computador portátil
- Tablet (iPod; galaxy tab, kindle, etc.)
- Cámara digital
- iPod / MP3 player

Elección de variables.

Se realizó el mismo procedimiento que en las secciones anteriores; escogiendo los valores más altos de las medianas de cada una de las variables que contiene el uso de dispositivos, este paso ayudara a elegir las variables de mayor relevancia. De las 7 variables que inicialmente que se contaba, al manejar este procedimiento de análisis de estadística de dependencia central, se obtuvo como resultado final 5 variables: Smartphone, teléfono móvil con acceso a Internet, teléfono móvil sin acceso a Internet, computador portátil y cámara digital.

Clasificación y verificación.

Para la clasificación por grupos se consideró realizar el análisis clúster k-medias, separándolos por sus características similares; se construyeron 3 clasificaciones para los clúster de 2, 3, y 4 (Anexo 13, 14 y 15). Finalmente para realizar la verificación del nivel de precisión de las

clasificaciones se aplicó el análisis discriminante utilizando como variables independientes las variables más relevantes y como variable dependiente el número de clúster. Se escogió la clasificación en dos grupos debido a que presenta mayor facilidad para interpretar, además da un nivel de precisión mayor que los clúster de 3 y 4. (Tabla 28).

Tabla 28. Porcentaje de precisión de la clasificación de dispositivos

Numero de grupos	Porcentaje de Exactitud
2	97,4 %
3	96,4 %
4	98,4 %

Fuente: Elaboración propia

Descripción de los grupos.

Para poder interpretar los grupos se hace énfasis en sus características o diferencias más comunes. Los centroides de las variables más significativas en cada clúster o conglomerados se pueden observar en la siguiente Figura 22:

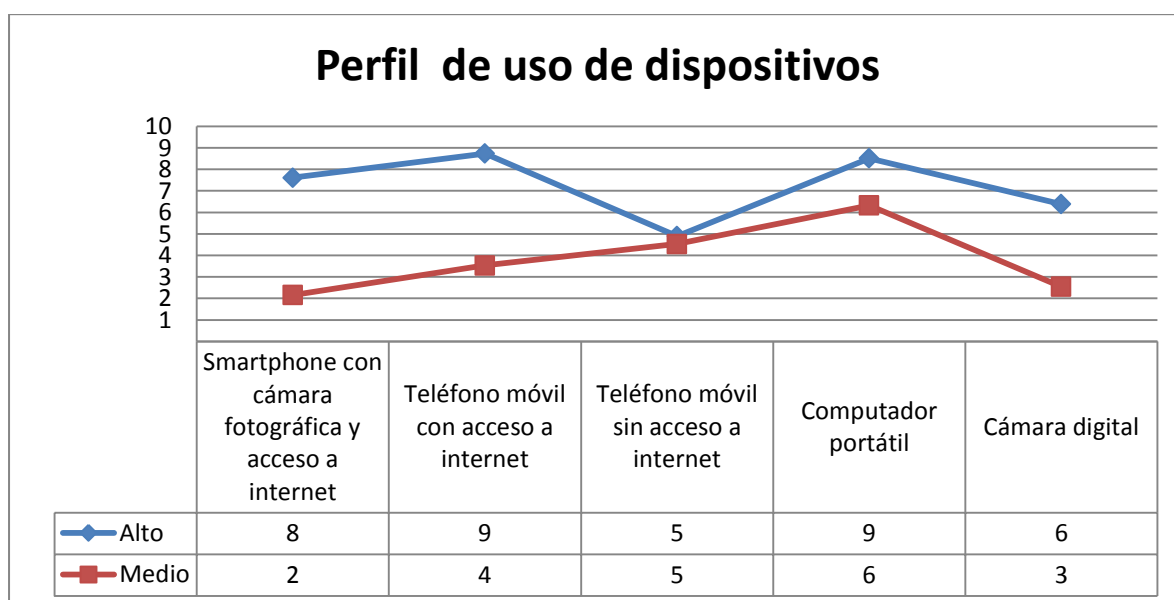


Figura 22. Perfil del uso de dispositivos

Fuente: Elaboración propia

El primer grupo, denominado “alto” tiene el 47,2% de estudiantes que se caracterizan por hacer uso de los dispositivos tanto de los modernos (el smartpocce, el telefono con acceso a

internet) como también de los tradicionales (teléfono móvil sin acceso a Internet, computador portátil, cámara digital) con valores altos en comparación con el grupo 2; sin embargo, se hace hincapié que en la variable de teléfono sin acceso a Internet hay un bajo uso de este dispositivo.

El segundo grupo denominado “medio” conforma el 52,8% de estudiantes, los mismos que tienen un perfil con las siguientes particularidades: estos usuarios se utilizan en mayor nivel el computador portátil, caso contrario, en menor valor el dispositivo smartphone, y la cámara digital, los demás factores como muestra la Figura 22 van creciendo su nivel de uso; sin embargo, su uso en general es medio en comparación con el grupo uno.

4.2. Fase de Evaluación e Interpretación de los datos.

La fase de Evaluación e interpretación de los datos es un método que permite evidenciar si las hipótesis propuestas son razonables y significativas. Para fundamentar con objetividad las hipótesis planteadas, necesitamos emplear algún proceso o método de minería de datos que se llegue a obtener un razonamiento objetivo para rechazar o aceptar las hipótesis planteadas; por tal razón, se utilizó el método estadístico del chi-cuadrado (χ^2), por ser una prueba que permite medir aspectos cualitativos y cuantitativos de las respuestas que el estudiante contestó en la investigación; además, este método estadístico ayuda a determinar si dos variables están relacionadas o no a nivel general, mientras que para predecir y explicar los resultados de variables categóricas en función de la variable independiente se utilizará la técnica de regresión logística multinomial (Odds Ratio), técnica que permite relacionar de forma más precisa y profunda la relación entre los niveles de la variable independiente con las categorías de la variable dependiente.

En este apartado procedemos a verificar las cuatro hipótesis propuestas para esta investigación, tanto en el aprendizaje, el entretenimiento como en el rendimiento académico.

4.2.1. Incidencia del nivel de ingresos en el uso del Internet para el aprendizaje.

Se contrasta la hipótesis 1 que sostiene que el nivel de ingresos determina como se utiliza Internet para el aprendizaje. En este estudio la variable dependiente cuenta con el clúster: perfil para actividades académicas y la variable independiente corresponde a la variable ingresos. La variable dependiente cuenta con las categorías: Medio =1, Interactivo=2, Pasivo =3; y la

variable ingresos tiene los siguientes valores: [ing=1,00]= hasta 350 dólares; [ing=2,00]= hasta 600 dólares; [ing=3,00]=hasta 1000; [ing=4,00]=hasta 1500; [ing=5,00]=más de 1500.

Para la comprobación del modelo en estudio se estableció que el nivel de ingresos esta relacionado significativamente con el uso de Internet en el aprendizaje académico, por lo que el valor del chi- cuadrado (Tabla7, p. 70) tiene un nivel de significancia menor a 0,05 ($\chi^2 = 23.476$; $p = 0,003$) con 8 grados de libertad. Considerando que la variable dependiente Perfil para actividades académicas “cluster3_acad” consta de tres niveles y es una variable categórica, se utiliza la regresión logística multinomial, ya que la misma permitirá observar el nivel de relación de una forma más precisa entre los grupos de manera iterativa.

Se realizó la verificación utilizando el valor de chi-cuadrado del logaritmo de la verosimilitud ($\chi^2 = 42,520$, $p= 0,003$), lo que determina un modelo significativo $p < 0,05$ (Anexo 16). Así, mismo a través de los estadísticos de Person y Deviance se obtuvo lo siguiente ($\chi^2 = 0,000$), para ambos casos; determinando que son valores significativos ($p= 0,000 < 0,05\%$), demostrando que no se ajusta los datos al modelo (Anexo 16). El valor de R^2 Nagelkerke da como resultado el 4,4% de la varianza explicada. Al utilizar la prueba de Wald se puede observar (Tabla 29); que en el modelo “Interactivo” y la variable ingresos tienen valores significativos solo en el nivel “[ing=3,00]” ($p=0,044 < 0,05\%$), mientras que en los demás niveles es todo lo contrario.

Tabla 29. Coeficientes del modelo de regresión logística (H1)

Número inicial de casosa	B	Error típ.	Wald	gl	Sig.	Exp(B)	Intervalo de confianza al 95% para Exp(B)		
							Límite inferior	Límite superior	
Interactivo (2)	Intersección	,693	,707	,961	1	,327			
	[ing=1,00]	,194	,753	,066	1	,797	1,214	,277	5,314
	[ing=2,00]	-,097	,747	,017	1	,896	,907	,210	3,920
	[ing=3,00]	-1,609	,800	4,047	1	,044	,200	,042	,959
	[ing=4,00]	-,182	,876	,043	1	,835	,833	,150	4,636
	[ing=5,00]	0	.	.	0
Pasivo (3)	Intersección	1,735	,626	7,673	1	,006			
	[ing=1,00]	,049	,669	,005	1	,941	1,050	,283	3,899
	[ing=2,00]	-,531	,664	,639	1	,424	,588	,160	2,160
	[ing=3,00]	-1,103	,673	2,685	1	,101	,332	,089	1,242
	[ing=4,00]	-,693	,786	,778	1	,378	,500	,107	2,333
	[ing=5,00]	0	.	.	0

a. La categoría de referencia es: 1 (Medio)

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

Fuente: Elaboración propia

La interpretación de resultados de la Tabla 29, permite observar que el Odds Ratio (OR) expresado como Exp (B), de corresponder al primer modelo de OR del perfil académico de Nivel Interactivo con respecto al Nivel Medio es 0,005 (1/0,200) veces menor cuando el estudiante pertenece al nivel de ingreso 3 (hasta 1000 dólares), en comparación con pertenecer al nivel 5 (OR = 0,200, (IC 95% 0,042 – 0,959), p= 0,000); no es significativo para los niveles de ingresos 1 (hasta 350 dólares), 2 (hasta 600 dólares), y 4 (hasta 1.500dólares), con respecto con pertenecer al nivel de ingreso 5 (más de 1.500 dólares)

En el segundo modelo del OR que corresponde al perfil académico de Nivel Pasivo con respecto al Nivel Medio, se observa que para todos los niveles de ingresos no existe una relación significativa ($p > 0,05\%$).

4.2.2. Incidencia del nivel de ingresos con el uso de Internet para entretenimiento.

Se comprueba la hipótesis 2 planteada en esta investigación, que sostiene que el nivel de ingresos determina como se utiliza Internet para el entretenimiento, tomándose en cuenta que la variable dependiente (perfil para entretenimiento) consta de tres niveles, se optó por diseñar el modelo de regresión logística multinomial, implicando examinar de forma más acertada las distintas posibilidades de asociación que existen. La variable dependiente cuenta con las categorías: inactivo =1; intermedio = 2; y dinámico = 3; y la variable ingresos consta de los siguientes valores [ing=1,00]= hasta 350 dólares; [ing=2,00]= hasta 600 dólares; [ing=3,00]=hasta 1000; [ing=4,00]=hasta 1500; [ing=5,00]=más de 1500.

En la verificación de la hipótesis 2 se establece que el nivel de ingresos no representa un asociación significativa con el uso de Internet para entretenimiento, ya que el valor del chi-cuadrado presentado en la Tabla 8 (p.71) del proyecto de investigación, tiene un nivel de significancia mayor a 0,05% ($X^2 = 11,285$, $p = 0,186$)

Aplicando el modelo regresión logística multinomial descrito anteriormente, se utilizó la prueba del logaritmo de la verosimilitud (-2 Log likelihood), resultando que el valor del chi-cuadrado es ($X^2 = 10,962$; $p = 0,204$), indicando que no existe un modelo significativo por su valor de significancia mayor a 0,05 (Anexo 17). Seguidamente verificando la segunda prueba de bondad de ajuste con los valores de Pearson ($X^2 = 0,000$) y Desviance ($X^2 = 0,000$), los dos valores resultantes tiene un nivel de significancia de $p = 0,000$; concluyendo que el modelo no se ajusta a

los datos (Anexo 17); en cuanto al nivel de varianza R^2 Nagelkerke alcanza un valor de 0,025 (2,5%).

Los resultados obtenidos al utilizar la prueba de test de Wald, tanto para los dos modelos en estudio, resultaron un nivel de significancia no aceptables ($p > 0.005$) para cada uno de los niveles de la variable ingreso (Tabla 30).

La deducción de los valores del Odds Ratio, presenta los resultados obtenidos de la variable dependiente (perfil para entretenimiento) en función de la variable independiente (ingresos), los mismos que permiten concluir que el nivel de ingresos no aporta valores significativos en este estudio investigativo; por tanto, no existe incidencia entre el nivel de ingresos y el uso de Internet para entretenimiento. Tabla 30.

Tabla 30. Coeficientes del modelo de regresión logística (H2)

Número inicial de casosa		B	Error típ.	Wald	gl	Sig.	Exp(B)	Intervalo de confianza al 95% para Exp(B)	
								Límite inferior	Límite superior
Dinámico (2)	Intersección	-,251	,504	,249	1	,618			
	[ing=1,00]	-,283	,548	,266	1	,606	,754	,257	2,207
	[ing=2,00]	-,682	,559	1,488	1	,222	,506	,169	1,512
	[ing=3,00]	-,116	,590	,039	1	,844	,890	,280	2,829
	[ing=4,00]	-1,483	,804	3,405	1	,065	,227	,047	1,097
	[ing=5,00]	0	.	.	0
Inactivo (3)	Intersección	,105	,459	,053	1	,819			
	[ing=1,00]	,488	,488	1,002	1	,317	1,629	,626	4,238
	[ing=2,00]	,178	,490	,132	1	,716	1,195	,458	3,121
	[ing=3,00]	,274	,525	,272	1	,602	1,315	,470	3,683
	[ing=4,00]	-,374	,589	,402	1	,526	,688	,217	2,183
	[ing=5,00]	0	.	.	0

a. La categoría de referencia es: 1.

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

Fuente: Elaboración propia

4.2.3. Incidencia del uso de la tecnología en el aprendizaje sobre el rendimiento académico.

En este apartado se verifica la hipótesis 3 que consiste en determinar si el uso de la tecnología en el aprendizaje incide en el rendimiento académico. Para su verificación, se analizó y comprobó los datos a nivel general utilizando el método del Chi- cuadrado, facilitando un

resultado entre estas dos variables categóricas; por lo que, se determinó que no existe una correlación significativa entre el uso de la tecnología en el aprendizaje sobre el rendimiento académico, por su alto valor en el nivel de significancia, sobrepasando el 0,05%. ($X^2 = 5.660$, $p= 0,059$) con 2 grados de libertad. Así mismo, se consideró aplicar la regresión logística binaria porque contamos con la variable dependiente (rendimiento_académico) con dos niveles, este modelo se realizó con el fin de verificar de una forma más exacta y profunda, la relación entre estas variables categóricas mediante el proceso iterativo que permite comprobar el nivel de significancia global y los coeficientes de estimación de las variables.

Se aplicó las pruebas de Ómnibus cuyos resultados muestran un valor Chi-cuadrado significativo ($X^2 = 3,988$, $p= 0,046$) con 1 grado de libertad, por ende se trata de una buena predicción del modelo. Al comprobar la bondad de ajuste mediante la prueba de Hosmer y Lemeshow presenta un valor de Chi-cuadrado ($X^2 = 2,666$; $p=0,103$) que no es significativo ($p>0,05$); por lo que podemos deducir que el modelo se ajusta a los datos, esto indica que al introducir la variable Independiente, no va mejora la predicción de la variable dependiente. El valor resultante de la prueba del R^2 de Nagelkerke en este modelo, alcanza un valor de varianza de 0,049 (4,9%) (Anexo 18).

En la comprobación de los coeficientes de la variables independientes se empleó la prueba de test de Wald como resultado obtuvimos valores no significativos ($p>0,05$). Así mismo, el valor del Odds Ratio no es significativo como lo indica la Tabla 31.

Tabla 31. Coeficientes del modelo de regresión logística binaria (H3)

		B	E.T.	Wald	gl	Sig.	Exp(B)
Paso 1 ^a	cluster3_acad	,752	,427	3,096	1	,078	2,122
	Constante	-5,190	1,190	19,009	1	,000	,006

a. Variable(s) introducida(s) en el paso 1: cluster3_acad.

Fuente: Elaboración propia

4.2.4. Incidencia del uso de la tecnología para entretenimiento en el rendimiento académico.

En esta sección se contrasta la hipótesis 4 planteada en esta investigación que sustenta que el uso de la tecnología para entretenimiento incide en el rendimiento académico. Para analizar

esta asociación se construye el modelo de regresión logística binomial. El perfil para entretenimiento (cluster3_entre) tiene tres categorías, y la variable dependiente rendimiento académico es una variable dicotómica: “0” para quienes aprueban el módulo y “1” para quienes no aprueban el módulo.

Al aplicar las pruebas para el modelo Omnibus como resultado se obtuvo que el valor de Chi-cuadrado no es significativo ($X^2 = 0,682$; $p = 0,409$), por lo que se deduce que el modelo no es predictivo (Anexo 19).

La comprobación de la bondad de ajuste se la realiza mediante la prueba de Hosmer y Lemeshow alcanzando un valor de Chi-cuadrado ($X^2 = 0,003$, $p = 0,953$) con 1 grado de libertad, obteniendo un modelo no significativo ($p > 0,05$); concluyendo que el modelo tiene un buen ajuste. El valor de nagelkerke indica que el modelo propuesto explica una varianza del 0,005. (Anexo 19), el cual es un valor muy bajo que indica la variable independiente no influyen en la variable dependiente.

El test de Wald se utilizó para la comprobación de los coeficientes de la variable independiente, en donde se alcanzó valores no significativos ($p > 0,05$). Por otro lado, el OR para este modelo no representa relaciones significativas, como resultado se rechaza la hipótesis 4 y podemos deducir que no existe incidencia entre el uso de las tecnologías para entretenimiento en el rendimiento académico en base al tema de investigación (Tabla 32).

Tabla 32. Coeficientes del modelo de regresión logística binaria (H4)

	B	E.T.	Wald	gl	Sig.	Exp(B)
Paso 1 ^a cluster3_entre	,221	,272	,659	1	,417	1,247
Constante	-3,732	,670	31,018	1	,000	,024

a. Variable(s) introducida(s) en el paso 1: cluster3_entre.

Fuente: Elaboración propia

CAPÍTULO 5: DISCUSIÓN DE RESULTADOS

En este apartado se da a conocer los hallazgos encontrados que se relacionan a las preguntas de investigación para alcanzar los objetivos planteados.

5.1. Internet en las preferencias de los estudiantes, en las actividades académicas y para entretenimiento.

La universidad Nacional de Loja es una universidad pública donde la información recolectada proporcionó una muestra estructurada de 504 estudiantes, conformado por el 58.7% de hombres y un 41.3% de mujeres, observando que existe mayor concentración en el género masculino que en el femenino.

Con respecto a la edad de los involucrados en la investigación está comprendida entre los 18 y 35 años; donde la mayor concentración de los datos (44.4%) están por encima de los 18 años de edad. Al relacionar la edad con las horas que se conecta a Internet el estudiante se verificó que existe una relación significativa entre estas dos variables ($p=0.000<0.05$), concluyendo que la edad influye en las horas de conexión a Internet; esto quiere decir, que los estudiantes de mayor edad tienden a emplear más horas de conexión a Internet que los de menor edad, lo cual se contrasta con lo que argumenta Poch (2009) que mayor edad, mayor es el número de horas de conexión a Internet”.

Se evidencia que el nivel de ingresos es un factor que brinda información importante para esta investigación, donde se pudo afirmar que existe mayor concentración en los niveles de ingresos familiares de 350 a 600 dólares mensuales, alcanzando el 72% de estudiantes inmersos en estos dos niveles, y tan solo el 5.2% tienen ingresos superiores a 1500 dólares por cada mes.

Se halló una relación significativa ($p<0.05\%$) entre los ingresos y el lugar de conexión a Internet; donde las familias con menores y mayores ingresos económicos prefieren conectarse y hacer uso de la tecnología desde su casa, pero las familias que cuentan con mayores posibilidades económicas prefieren accesos selectivos como los cyber, desde su trabajo, la universidad o desde un teléfono móvil; sin embargo, los estudiantes que utilizan estas opciones se encuentran introducidos en un porcentaje muy bajo ($<10\%$). Según Jansen y Fellow (2010) comentan que los hogares con niveles de ingresos más alto, tienen mayores posibilidades de conectarse desde su casa; mientras que Serrano y Martinez (2003), argumentan que “el ingreso determina los bienes y servicios que podemos adquirir, por lo que si el ingreso es bajo, solo se

tendrá lo necesario para vivir y el acceso a las TIC no es prioritario y queda en último término” (p. 114).

Los niveles de ingresos influyen significativamente en los días que el estudiante se conecta a Internet a la semana, como también en los niveles de conocimiento en el manejo del Internet, demostrando que a pesar de contar con ingresos familiares bajos, los estudiantes de la UNL tienen un nivel de conocimiento alto, esto es en referencia de una escala de 1 a 10, donde el uno representa la inexperiencia y 10 representa el nivel de conocimiento máximo en el manejo del Internet. Lo que quiere decir que tan solo el 55.6% de los estudiantes con ingresos de hasta 350 dólares mensuales tienen una frecuencia de utilizar el Internet los 7 días de la semana alcanzando un nivel 8 de conocimiento en el manejo del Internet.

Otro hallazgo encontrado durante la investigación, es el efecto que causa los ingresos en las horas que el estudiante se conecta a Internet y los años de experiencia en el mismo, resultando una asociación significativa $p < 0.05\%$; en donde, los años de experiencia y las horas que se conectan a Internet dependen de los ingresos familiares del estudiante. Con los resultados obtenidos los estudiantes que presentan niveles de ingresos bajos utilizan el Internet de una a tres horas diarias logrando una experiencia de por lo menos de cuatro a seis años, esto difiere con el análisis de Taylor et al. (2003), quienes argumenta que las personas que alcanzan a obtener niveles de ingresos altos tienen mayor posibilidad de estar más tiempo conectados a Internet, alcanzando niveles más altos en conocimientos y años de experiencias.

Referente el uso de los dispositivos tecnológicos, se determinaron dos grupos:

- El primer grupo denominado perfil alto consta de 47,2% de los estudiantes, su característica principal es el manejo con un valor alto de los dispositivos modernos como de los tradicionales; sin embargo, posee un valor bajo en el uso del teléfono sin acceso a Internet.
- El segundo grupo denominado perfil medio, tiene el 52,8% de estudiantes, caracterizándose por tener un uso medio en sus variables, aunque tiene una particularidad que usan en mayor nivel el computador portátil en comparación con el smartphone y la cámara digital.

Por tanto, es necesario destacar que en los dos perfiles obtenidos: medio y alto, el uso de los dispositivos que se manejan pueden ayudar a que los estudiantes prueben las nuevas tendencias tecnológicas que ofrece el Internet, mejorando sus conocimientos, habilidades y destrezas; finiquitando que el uso de dispositivos dependen significativamente de los ingresos familiares del estudiante.

En el trabajo investigativo es necesario señalar los siguientes resultados obtenidos en cada una de las variables contenidas en el uso del Internet en las actividades académicas; sin embargo, se debe tener en cuenta que los estudiantes pertenecen a la modalidad presencial, donde la plataforma virtual es una opción.

- Referente a la plataforma virtual: se concluye que la mayor parte de los estudiantes al menos ingresan una vez a la semana a la plataforma virtual.
- El 52% de los estudiantes realizan de una a cinco consultas a sus profesores por mes, mientras que el 14.7% no lo hacen, esto puede dar a entender que estos estudiantes pueden hacer sus consultas sin el uso de la tecnología; es decir, personalmente o simplemente no se interesan en ese aspecto.
- Con respecto a las consultas realizadas a los compañeros, se evidencia que se realizan de una a cinco consultas por mes para temas académicos
- Los estudiantes que utilizan esta opción en sus actividades académicas tienen un promedio alrededor de tres descargas de recursos educativos al mes desde la plataforma virtual.
- La mayor proporción de estudiantes observan un máximo de 6 videos académicos al mes, mientras que el 8.5% no lo hacen.
- Se pudo observar que existe poco uso de los foros virtuales y tweets al mes, en donde la mayor parte de los estudiantes no utilizan estas opciones tecnológicas.
- El uso del chat en actividades académicas es de una a cuatro horas al mes.
- Los estudiantes emplean un promedio de 16,76 horas en la búsqueda de información académica de manera mensual con una tendencia a variar superior o inferior a 13,8 en el uso de otros temas.
- Se pudo observar que los estudiantes utilizan muy poco la biblioteca virtual, con promedio de 4.2 horas al mes.

Se llega a comprobar con los resultados obtenidos anteriormente que los estudiantes que

pertenecen a la UNL, emplean su tiempo y dedicación en hacer uso de la tecnología para actividades académicas; en donde se resalta la interacción con sus compañeros de clase por medio del chat académico, como en la búsqueda de información académica.

Por otro lado se analizan los resultados obtenidos en el uso de la tecnología para entretenimiento:

- Se evidencia que la mayor parte de los estudiantes (46.6%) emplean sus horas de chat por diversión entre uno a cinco horas semanales.
- El 36.3% de los estudiantes utilizan las redes sociales como punto de diversión; mientras que tan solo el 1,4% no lo hacen.
- Alrededor del 59.3% de los estudiantes no utilizan los juegos en línea, pero existe un porcentaje pequeño que utilizan por lo menos una hora a la semana,
- Con respecto a la descarga de música, videos y programas para entretenimiento, se observó que los estudiantes dedican de una a tres horas por semana para hacer uso de estas opciones tecnológicas.
- Los estudiantes cuentan con un promedio de 6,82 videos vistos en youtube por diversión.

Al verificar los resultados expuestos se pudo evidenciar que los estudiantes emplean su mayor tiempo en las redes sociales para su entretenimiento, lo cual se determina que esta variable predomina de las demás; así mismo, toda la población estudiantil utilizan las demás opciones que ofrece la tecnología para su vida social pero en menor proporción.

5.2. Efecto del nivel de ingresos sobre los usos de Internet en actividades académicas.

Por lo expuesto en el capítulo anterior se ha llegado a unificar el análisis de datos en la siguiente Tabla (referente a la primera hipótesis), indicando las variables escogidas por sus valores más significativos, usando la técnica de minería como la clusterización que permitirá agrupar a los estudiantes acorde a sus características similares; así mismo, da a conocer los resultados de la clusterización escogida de tres grupos, con su respectivo nivel de discriminación, llegando a determinar si existe incidencia entre la variable independiente(VI) sobre la dependiente(VD).

- **H1:** El nivel de ingresos determina como se utiliza Internet para las actividades

académicas.

- **H0:** El nivel de ingresos no determina como se utiliza Internet para las actividades académicas.
- **V.I.:** ingresos
- **V.D.:** perfil para actividades académicas

Tabla 33. Análisis de los datos de la H1

Variables	Resultado de la clusterización	Porcentaje del nivel de discriminación	Comprobación de hipótesis (Regresión)	
V.D. a) descargas de recursos educativos b) consultas a los profesores c) consultas a los compañeros d) videos vistos en youtube e) búsqueda de información	perfil para actividades académicas conformado: grupo1: medio grupo2: interactivo grupo3: bajo	94.8%	Valor Chi cuadrado de Pearson	$X^2=23.476$; $p=0.003$
			Valor Chi cuadrado del logaritmo de verosimilitud	$X^2=42.520$; $p=0.003$ (valor significativo: modelo significativo)
			Valor de Pearson y Deviance	$X^2=0.000$; $p=0.000$ (valor significativo: no se ajusta a los datos)
			Valor R² Nagelkerke	4.4% de la varianza (valor muy bajo)
V.I. ingresos	N/A	N/A	Prueba de Wald	No todas las variables alcanzan un valor significativo ($p<0.05$), excepto en el nivel 3 del grupo interactivo
			Odd Ratio (OR)	Existe una baja incidencia ($p=0,044$) en el nivel 3 de ingresos perteneciente al grupo interactivo respecto al grupo medio. Con respecto a los otros niveles del grupo interactivo y del grupo bajo no existe incidencia alguna.

Fuente: Elaboración propia

Esto comprueba que si existe incidencia en la hipótesis 1, por lo que se concluye que el nivel de ingresos incide en las actividades académicas de los estudiantes, respaldando estos resultados con las conclusiones que llegan algunos autores Zillien y Hargittai (2009); Ruiz y Ortiz (2014).

5.3. Efecto del nivel de ingresos sobre los usos de Internet para el entretenimiento.

- **H₀:** El nivel de ingresos determina como se utiliza Internet para entretenimiento.
- **H₁:** El nivel de ingresos no determina como se utiliza Internet para entretenimiento.
- **V.I.:** ingresos
- **V.D.:** perfil para entretenimiento

Tabla 34. Análisis de los datos de la H2

Variables	Resultado de la clusterización	Porcentaje del nivel de discriminación	Comprobación de hipótesis (Regresión)	
			Valor Chi cuadrado de Pearson	Valor Chi cuadrado del logaritmo de verosimilitud
V.D. a) Horas que se chatea por diversión b) uso de redes sociales c) videos para entretenimiento vistos en youtube	Perfil para entretenimiento conformado: grupo1: intermedio grupo2: dinámico grupo3: inactivo	98%	Valor Chi cuadrado de Pearson	$X^2=11,285$; $p=0.186$ (el valor no es significativo)
			Valor Chi cuadrado del logaritmo de verosimilitud	$X^2=10.962$; $p=0.204$ (valor no significativo: modelo no significativo)
			Valor de Pearson y Deviance	$X^2=0.000$; $p=0.000$ (valor significativo: no se ajusta a los datos)
			Valor R^2 Nagelkerke	2.5% de la varianza (valor muy bajo)
V.I. ingresos	N/A	N/A	Prueba de Wald	Todas las variables no alcanzan un valor significativo ($p<0.05$)
			Odd Ratio (OR)	No existe incidencia alguna.

Fuente: Elaboración propia

La Tabla 34 presenta el análisis que se realizó a las variables inmersas en la hipótesis 2, por lo que al analizar el efecto que causa el nivel de ingreso sobre los usos de Internet para el entretenimiento se ha llegado a la conclusión que en estas dos variables en estudio no existe incidencia alguna, rechazándose la hipótesis nula y aceptando la hipótesis alternativa. Esto coincide con lo que menciona Ruiz y Ortiz (2014) quien sostiene que el Internet se utiliza principalmente como medio de entretenimiento, observándose un mayor uso como medio de comunicación (email, chat, entre otros) en la población con educación superior y nivel de posgrado “impactando de manera positiva al ingreso y gasto total de los hogares” (p. 371).

5.4. Efecto de los usos de Internet en actividades académicas sobre el rendimiento académico

- **H₀:** El uso de la tecnología en el aprendizaje incide en el rendimiento académico
- **H₁:** El uso de la tecnología en el aprendizaje no incide en el rendimiento académico
- **V.I.:** perfil para actividades académicas
- **V.D.:** rendimiento académico

A continuación, se presentan los resultados obtenidos durante la investigación concerniente a la hipótesis tres.

Tabla 35. Análisis de los datos de la H3

Variables		Comprobación de hipótesis (Regresión)	
V.I.		Valor Chi cuadrado de Pearson	$X^2=5.660$; $p=0.059$ (el valor no es significativo)
Perfil para actividades académicas	grupo1: medio grupo2: interactivo grupo3: bajo	Valor Ómnibus	$X^2=3.988$; $p=0.046$ (valor significativo: modelo significativo)
		Valor de Hosmer y Lemeshow	$X^2=2.666$; $p=0.103$ (valor no significativo: modelo se ajusta a los datos)
		Valor R^2 Nagelkerke	4.9% de la varianza (valor muy bajo)
V.D.		Prueba de Wald	Las variables no alcanzan un valor significativo ($p<0.05$)
Rendimiento académico	0: aprobado 1: no aprobado	Odd Ratio (OR)	No existe incidencia alguna.

Fuente: Elaboración propia

Al realizar la relación de los usos de Internet en las actividades académicas sobre el rendimiento académico se comprobó que no existe incidencia entre estas dos variables coincidiendo lo que sostiene según sus estudios Ayala (2007); además según Luna (2012) sus resultados concuerdan con las conclusiones obtenidas en la investigación llegándose a determinar que el uso de Internet no incide en el rendimiento académico del estudiante, sugiriendo replantear “el uso de tecnología como parte de las actividades de aprendizaje buscando un efecto positivo especialmente en la calificación que logra el estudiante”. Otros autores Raines (2012); Luaran, J. E., Binti, F., Mohd, K., y Nadzri, F. Luaran (2011) también se suman a las conclusiones conseguidas.

Estos resultados obtenidos va en contra de las conclusiones que obtiene en su investigación Torres(2012), donde sostiene que los usos del Internet presentan efectos positivos para el rendimiento académico; así mismo, lo respaldan con los mismos resultados otros autores (Ayala, 2007; Kuh y Hu, 2001). Según Castaño-Muñoz (2010), afirma que la “introducción de las herramientas tecnológicas en el aprendizaje sólo es eficaz para mejorar el rendimiento académico cuando se explota su capacidad de cambio” (p. 7).

Gil-Flores (2009), quien demostró que la frecuencia al utilizar el computador desde la casa, incide en el éxito académico del estudiante en cuanto a sus calificaciones; es decir, que existe una relación entre el uso del computador y el rendimiento académico. Así mismo, este estudio coincide con la investigación realizada por López Goycochea (2005), en donde afirma que “el uso adecuado de la computadora como herramienta de aprendizaje, permite mejorar el rendimiento académico de los alumnos en forma significativa y también en el logro de sus competencias”. Esto difiere con el estudio que realizó Warschauer (2008) sobre las escuelas primarias, donde concluye no existe una incidencia del uso de computadores portátiles en la mejora del rendimiento académico. De igual forma, Wittwer, J. y Senkbeil (2005) concluyen que

no hay incidencia alguna entre el uso de computadoras y el desempeño académico del estudiante.

Duart, Gil, Puyol, y Castaño (2008) explican en el análisis de las universidades catalanas, refiriéndose a los créditos aprobados del estudiante, en donde se consideró el número de aprobación de créditos con respecto al número de créditos que el estudiante se ha matriculado, consiguiendo un indicador del logro académico, el cual determina que las mujeres aprueban sus créditos, mientras que los hombres tienden a reprobar más, además se llegó a la conclusión que los estudiantes menores de 25 años presentan un mayor porcentaje de créditos aprobados obteniendo un alto rendimiento. Por otro lado, Conde, Ruiz, y Torres-Lana (2010) sostienen que durante su análisis e investigación sobre el rendimiento académico refiriéndose a las calificaciones de los estudiantes, no encuentra relación entre el tiempo de uso en Internet con el rendimiento académico, mientras que si obtuvo evidencia sobre el tiempo que se conecta el estudiantes con las asignaturas reprobadas.

5.5. Efecto de los usos de Internet en entretenimiento sobre el rendimiento académico

- **H₀**: El uso de la tecnología para el entretenimiento incide en el rendimiento académico
- **H₁**: El uso de la tecnología para el entretenimiento no incide en el rendimiento académico
- **V.I.**: perfil para entretenimiento
- **V.D.**: rendimiento académico

Tabla 36. Análisis de los datos de la H4

Variables		Comprobación de hipótesis (Regresión)	
V.I.		Valor Chi cuadrado de Pearson	$\chi^2=0.674$; $p=0.714$ (el valor no es significativo)
Perfil para entretenimiento	grupo1: intermedio grupo2: dinámico grupo3: inactivo	Valor Ómnibus	$\chi^2=0.682$; $p=0.409$ (valor no significativo: modelo no ajustado)
		Valor de Hosmer y Lemeshow	$\chi^2=0.003$; $p=0.953$ (valor no significativo: modelo se ajusta a los datos)
		Valor R ² Nagelkerke	0,5% de la varianza (valor excesivamente bajo)
V.D.		Prueba de Wald	Las variables no alcanzan un valor significativo ($p<0.05$)
Rendimiento académico	0: aprobado 1: no aprobado	Odd Ratio (OR)	No existe incidencia alguna.

Fuente: Elaboración propia

En este apartado los resultados obtenidos no resultó favorable (Tabla 36), ya que se pudo comprobar que al construir el modelo de predicción con regresión logística binaria, se verificó

que la probabilidad de acierto no varía; es decir, que el estudiante al hacer uso de las tecnologías del Internet con propósitos para entretenimiento y diversión no influyen en su rendimiento académico, rechazándose la hipótesis nula y aceptando la hipótesis alternativa; por tanto, el estudiante para alcanzar un buen rendimiento académico favorable, lo puede realizar independientemente si utilizan en mayor o menor proporción el uso de la tecnología para su entretenimiento; se une a las conclusiones que han llegado Frangos, Frangos, y Kiohos (2010), donde explican que el uso del Internet para entretenimiento no influyen en el rendimiento académico del estudiante.

Estos resultados difieren con las conclusiones realizadas por Castaño-Muñoz (2010), que sostiene que en los países desarrollados el uso del Internet si afecta en el rendimiento académico obteniéndose efectos negativos al hacer uso del Internet con finalidades de diversión. Además, apoyando a las conclusiones anteriores según las investigaciones de Torres (2012) determinó que el uso del Internet de forma mínima incide en el rendimiento académico del estudiante; sin embargo, propone para investigaciones futuras se amplíe la investigación con el fin de aclarar de manera más favorable con los resultados de la investigación realizada, obteniéndose información adicional sobre las actividades para entretenimiento que usan los estudiantes.

Otra investigación realizada por Salazar (2013) realizada a estudiantes secundarios del colegio Menor Indoamérica, de la ciudad de Ambato, sostiene que las redes sociales de Internet inciden en el rendimiento académico de manera negativa en los estudiantes; en donde, las redes sociales son utilizadas en un intervalo de 3 a 5 horas diarias para entretenimiento descuidado las actividades académicas. En el mismo estudio los estudiantes consideran conveniente el uso de las redes sociales como facebook para crear grupos de estudio que permitan mejorar su rendimiento académico.

CAPÍTULO 6: CONCLUSIONES Y RECOMENDACIONES

6.1. Conclusiones.

En este apartado se recoge todas las conclusiones que se ha llegado en el transcurso de la investigación en base a los modelos diseñados.

- El 39.1% de los estudiantes tienen ingresos familiares de 350 dólares mensuales; sin embargo, adquieren una conexión desde su hogar, haciendo uso del Internet los 7 días de la semana con un promedio de una a tres horas al día.
- Se demostró que el nivel de ingresos familiar de los estudiantes de la UNL, tiene una baja incidencia sobre el uso del Internet en las actividades académicas, puesto que el 58.1% de las variables de estudio (descarga de recursos educativos, consultas a los profesores, consultas a los compañeros, videos vistos en youtube y búsqueda de información académica) demuestran que existen poca iteración en la plataforma virtual y bajo uso de actividades académicas en la Web 2.0
- Se comprobó que no existe una relación significativa entre los niveles de ingresos familiares y el uso del Internet para entretenimiento; entendiéndose al construir el modelo de predicción mediante regresión logística multinomial, este no se ajustó a los datos; además el valor del Chi-cuadrado no es significativo ($p=0.188$; $p>0.5\%$).
- Se evidenció que el uso del Internet tanto en las actividades académicas como para entretenimiento no inciden significativamente sobre el rendimiento académico, ya que los modelos de predicción construidos con la regresión logística binomial, dan como resultado que la probabilidad de acierto no varía al introducir la variable independiente sobre la dependiente.
- Al aplicar una de las técnicas de minería de datos como la factorización para la agrupación de los datos, se llegó a determinar que las variables de estudio no cumplieron con las condiciones que plantea esta técnica; por lo que se optó por emplear otra técnica descriptiva como la clusterización con la finalidad de determinar los conglomerados en el uso del Internet para actividades académicas como de entretenimiento, se comprobó que esta técnica fue la más viable y que mejor se adaptó a los datos.

- El algoritmo K-medias permitió agrupar a los estudiantes de acuerdo a sus características similares, facilitando su descripción e interpretación de los datos; generando una nueva variable que permitirá relacionarlas con las variables presentes en las hipótesis planteadas.
- La elección del modelo de predicción depende del problema planteado o de los datos objetos de estudio, como por ejemplo, para las dos primeras hipótesis se utilizó la regresión logística multinomial, puesto que contaba con una variable dependiente de tres categorías; mientras que para dar contestación a las dos últimas hipótesis se utilizó la regresión logística binaria, ya que la variable dependiente contaba con dos grupos.

6.2. Recomendaciones.

En esta sección se presentan las respectivas recomendaciones, información que ayudarán a futuras investigaciones.

- Es necesario concientizar a los estudiantes de los esfuerzos económicos que realizan sus familias con el propósito de brindarles la utilización del uso del Internet para sus diferentes actividades, pese que cuentan con niveles de ingresos económicos bajos.
- Si bien es cierto que el nivel de ingresos familiar incide en las actividades académicas, se observa que existe una relación mínima, en donde se deberá determinar que otros factores puede afectar su uso; como también, que otros elementos pueden determinar una fuerte incidencia de los ingresos familiares sobre el uso del Internet para actividades académicas como de entretenimiento.
- Es necesario identificar claramente las variables que dependen la una de la otra; con la finalidad de obtener resultados que aporten a los objetivos de la investigación determinando efecto real de las herramientas tecnológicas dentro del sistema educativo.
- Se recomienda utilizar varios algoritmos dentro de la técnicas de minería de datos como: K-medias, factorización, clusterización de manera que cada uno de un conocimiento basado en sus características particulares y que permitirá determinar las grupos de variables más significativas que se debe poner atención a la hora de mejorar y poder tomar decisiones.

- Sería conveniente evaluar y reestructurar de forma más minuciosa la encuesta planteada para los estudiantes; con la finalidad de obtener otras variables que aporten a la investigación y descartando las que no brinda información relevante.
- Es preciso extender y profundizar esta investigación con el fin de contar con una mayor información que permita aprovechar de mejor manera el uso del Internet, optando por un uso ético de las mismas, para que las universidades pueden implementar o hacer una evaluación exhaustiva de las nuevas tecnologías que brinda el Internet de manera eficiente que permita una reestructuración en su infraestructura tecnológica.

REFERENCIAS

- Acid, S., Marin, N., Medina, J., Pons, O., y Vila, A. (2005). *Introducción a las bases de datos: el modelo relacional*. (Editorial Paraninfo, Ed.) (reimpresa.).
- Aguayo, M. (2007). Cómo hacer una Regresión Logística con SPSS «paso a paso» (I). *DocuWeb-fabis. Huelva: Fundación Andaluza Beturia ...*, (1), 1-16. Recuperado de: http://www.fabis.org/html/archivos/docuweb/Regres_log_1r.pdf
- Aladi. (2003). La brecha digital y sus repercusiones en los países miembros de la ALADI. *Asociación Latinoamericana de Integración*.
- Aldehuela, M. (2005). *Análisis comparativo entre métodos estadísticos y de minería de datos*. Universidad Pontificia Comillas. Recuperado de: <http://www.iit.upcomillas.es/pfc/resumenes/42b98a68079d0.pdf>
- Álvarez. (2011). *Internet y derechos fundamentales*. (Purrúa, Ed.). México.
- Álvarez, R. (2007). *Estadística multivariante y no paramétrica con SPSS*: (Ediciones Díaz de Santos, Ed.) (Ilustrada.). Madrid.España.
- Angeles, M., y Santillán, A. (1988). Minería de datos: concepto, características, estructura y aplicaciones. Recuperado de: <http://www.ejournal.unam.mx/rca/190/RCA19007.pdf>
- Arnal, J., del Rincón, D., y Latorre, A. (1992). *Investigación educativa. Fundamentos y metodología*. (Labor, Ed.). Barcelona-España.
- Ayala. (2007). Relación entre el uso de Internet y el Logro Académico.
- Belloch, C. (2012). *Las Tecnologías de la Información y Comunicación en el aprendizaje*. Departamento de Métodos de Investigación y Diagnóstico en Educación. Universidad de Valencia. Recuperado de: <http://www.uv.es/bellohc/pedagogia/EVA1.pdf>
- Bernal, E. (2013). *Bioestadística Básica para Investigadores con SPSS*. (Bubok, Ed.). 27 de diciembre.
- Berzal, F. (2005). Metodos de agrupamiento: clustering, 3. Recuperado de: <http://elvex.ugr.es/doc/proyecto/cap8.pdf>
- Calcagno, S., y Hernández, I. (2003). *Los pueblos indígenas y la sociedad de la información en América Latina y el Caribe. Un marco para la acción*”. Recuperado de: <http://lanic.utexas.edu/project/etext/llilas/claspo/workingpapers/indigenas.pdf>
- Camacho, K. (2007). Tipos de Brecha Digital. Recuperado de: <https://jbalcarcel.wordpress.com/2008/05/27/tipos-de-brecha-digital/>
- Canela, M., Lora, M., y Estrella. (2011). Cómo hacer una Regresión Logística binaria «paso a paso» (II): análisis multivariante. *Docuweb FABIS*, 34.

- Castañeda, L. (2010). *Aprendiendo con Redes Sociales. Tejidos educativos para nuevos entornos*. (S. L. E. Editorial MAD, Ed.) (Junio,2010 ed.). Sevilla. España.
- Castaño-Muñoz, J. (2010). La desigualdad digital entre los alumnos universitarios de los países desarrollados y su relación con el rendimiento académico. *Revista de Universidad y Sociedad del Conocimiento, Universitat Oberta de Catalunya España*, 7(1), 1-11.
Recuperado de: <http://redalyc.uaemex.mx/src/inicio/ArtPdfRed.jsp?iCve=78012953012>
- Castells, M. (2001). Internet y La Sociedad Red. En *Lección inaugural del programa de doctorado sobre la sociedad de la información y el conocimiento* (p. 5). Universitat Oberta de Catalunya (UOC),.
- Chakraborty, J., y Bosman, M. (2005). Measuring the Digital Divide in the United States: Race, Income, and Personal Computer Ownership. *The Professional Geographer*, 57(3).
- Chávez, D., Miranda, I., Varela, M., y Fernández, L. (2010). Utilización del análisis de cluster con variables mixtas en la selección de genotipos de maíz (*Zea mays*). *Revista Investigación Operacional*, 30(3), 209-216. Recuperado de: rev-inv-ope.univ-paris1.fr/files/31310/31310-03.pdf
- Concorbado, M., y Gómez, D. (1994). Base de datos Documentales. *Educación y Biblioteca*, 45, 64. Recuperado de:
http://gredos.usal.es/jspui/bitstream/10366/112749/1/EB06_N045_P64-69.pdf
- Conde, E., Ruiz, C., y Torres-Lana, E. (2010). Relación entre el uso de Internet y el rendimiento académico en una muestra de adolescentes canarios. *Universidad de La Laguna*.
- Conti, D., y Martínez de Pinzón, J. (2007). *Reglas de Asociación en Series Temporales: panorama referencial y tendencias*. Logroño. Venezuela. Recuperado de:
<http://www.lsi.us.es/redmidas/CEDI07/%5B22%5D.pdf>
- Cooper, M. (2005). Inequality in the digital society: why the digital divides deserves all the attention it gets.
- De Haro, J. (2010). *Redes sociales para la educación*. (Anaya Multimedia, Ed.). Madrid.España.
- De la Fuente, S. (2011a). Análisis De Conglomerados. *Fac. Ciencias Económicas y Empresariales*, 55. Recuperado de:
<http://www.fuenterrebollo.com/Economicas/ECONOMETRIA/SEGMENTACION/CONGLOMERADOS/conglomerados.pdf>
- De la Fuente, S. (2011b). Análisis Discriminante. *Fac. Ciencias Económicas y Empresariales. UAM*.
- De la Fuente, S. (2011c). Análisis factorial. *Fac. Ciencias Económicas y Empresariales. UAM*. Recuperado de:
<http://www.fuenterrebollo.com/Economicas/ECONOMETRIA/MULTIVARIANTE/FACTORIAL/analisis-factorial.pdf>

- De la Fuente, S. (2011d). Regresión Logística. *Fac. Ciencias Económicas y Empresariales. UAM.*
- Díaz De Rada, J. V. (1998). Diseño de tipologías de consumidores mediante la utilización conjunta del Análisis Cluster y otras técnicas multivariantes. *Revista Española de Economía Agraria*, (182), 75–104. Recuperado de: <http://dialnet.unirioja.es/servlet/articulo?codigo=721363>
- Díaz De Rada, J. V. (2002). *Técnicas de análisis multivariante para investigación social y comercial.* (RA-MA S.A., Ed.) *International Journal. Madrid: RAMA.* Madrid.España: 14 de Junio de 2002.
- Dimaggio, P., Hargittai, E., Neuman, R., y Robinson, J. (2004). Fromun equal access to differentiated use: A literaturereviewand agenda forresearchon digital inequality. *In K. (Russell S. F. Neckerman (Ed.), Social inequality*, 355–400.
- Duart, Gil, M., Puyol, M., y Castaño, J. (2008). La universidad en la sociedad red. *Barcelona: Ariel.*
- EduTIC Ecuador. (2014). Noticias Ecuador - Ecuador en vivo - 4G LTE y Small Cells garantizan un 100% de rendimiento, disponibilidad y capacidad de conexión a internet | Tecnología. Recuperado de: <http://www.edutic.ec/home/blog-edutic-ecuador>
- Fayyad, U., Piatetsky-Shapiro, G., y Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. *American Association for Artificial Intelligence*, 1-34. Recuperado de: <http://www.csd.uwo.ca/faculty/ling/cs435/fayyad.pdf>
- Floridi, L. (2007). A look into the future impact of ICT in our lifes. *La Sociedad de la Información: An International*, 23(1). doi:10.1080 / 01972240601059094
- Foro Económico Mundial. (2014). Networked Readness Index.
- Frangos, C., Frangos, C., y Kiohos, a. (2010). Internet Addiction among Greek University Students: Demographic Associations with the Phenomenon, Using the Greek Version of Young's Internet Addiction Test, 1-35. Recuperado de: <http://discovery.ucl.ac.uk/20369/>
- Fundacion Chandra. (2007). Herramientas y buenas prácticas para las organizaciones no lucrativas en el uso de las TICs. *Laboratorio de innovación social.*
- Gil-Flores, J. (2009). Computer use andstudents' academicachievement. *Research, ReflectionsandInnovations in Integrating ICT in Education*, 1291-1295.
- Gómez, M., Roses, S., y Farías, P. (2012). El uso académico de las redes sociales en universitarios. *Revista Científica de Educomunicación.* Recuperado de: <http://www.revistacomunicar.com/index.php?contenido=detallesy numero=38yarticulo=38-2012-16>

- González, A., Soltero, M. de L., y González, E. (2012). *Efectos educativos de las redes sociales en alumnos del Nivel Medio Superior en Chihuahua*. Mexico. Recuperado de: [http://www.fca.uach.mx/apcam/2014/04/05/Ponencia 65-UACH.pdf](http://www.fca.uach.mx/apcam/2014/04/05/Ponencia%2065-UACH.pdf)
- Graham, R. (2008). The Stylization of Internet Life?: Predictors of Internet Leisure Patterns Using Digital Inequality and Status Group Perspectives, *13*(5).
- Gutiérrez, L., y Gamboa, L. (2010). Determinants of ICT Usage among Low-Income Groups in Colombia, Mexico and Peru. The information society. *La Sociedad de la Información: An International Journal*, *26*(5), 346–363. doi:10.1080 / 01972243.2010.511559
- Halpern, D. (2014). Consecuencias sociales uso de TICs en escolares. Chile. Recuperado de: <http://tren-digital.cl/estudio/usoticasescolares/>
- Hargittai, E. (2010). Digital Natives 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
- Hasperué, W. (2012). *Extracción de Conocimiento en grandes bases de datos utilizando estrategias adaptivas*. Universidad Nacional de la Plata.
- Hernández, J., Ramirez, J., y Ferri, C. (2004). *Introducción a la Minería de datos*. (Prentice Hall, Ed.) (Reimpresión). Madrid. España: 2007.
- Hernández-Rodríguez, O. (1998). *Temas de Análisis Estadístico Multivariado*. (Editorial Universidad de Costa Rica, Ed.) (1998.^a ed.). San José. Costa Rica.
- Herrero, J. (2010). El Análisis Factorial Confirmatorio en el estudio de la Estructura y Estabilidad de los Instrumentos de Evaluación: Un ejemplo con el Cuestionario de Autoestima CA-14. *Intervención Psicosocial*, *19*(3), 289-300. doi:10.5093/in2010v19n3a9
- Hilbert, M. (2010). When is Cheap, Cheap Enough to Bridge the Digital Divide? Modeling Income Related Structural Challenges of Technology Diffusion in Latin America. *World Development*, *38*(5), 756-770. doi:doi:10.1016/j.worlddev.2009.11.019
- Hosmer, D. W., y Lemeshow, S. (2004). *Applied logistic regression*. (John Wiley y Sons, Ed.).
- Howard, P. E. N., Rainie, L., y Jones, S. (2001). The Impact of a Diffusing Technology. *American Behavioral Scientist*, 382-404.
- INEC. (2013). Tecnologías de la Información y Comunicaciones (TIC'S). *Estadísticas Sociales INEC 2013*, 1-33. doi:10.1109/ColombianCC.2013.6637543
- Iske, S., Klein, A., y Kutscher, N. (2005). Differences in Internet Usage -- Social Inequality and Informal Education. *Social Work y Society*, *3*(2), 215-223.
- Jansen, J., y Fellow, S. (2010). Use of the internet in higher- income households. Washington. Recuperado de: <http://it.ejo.ch/wp-content/uploads/PIP-Better-off-households-final2.pdf>

- Kuh, G., y Hu, S. (2001). The Relationships Between Computer and Information Technology Use, Selected Learning and Personal Development Outcomes, and Other College Experiences. *Journal of College Student Development*, 42, 217–232. Recuperado de: <http://www4.ncsu.edu/~ladare/eac595/readings/ku-hu.pdf>
- López Goycochea, J. (2005). *Relación entre el Rendimiento Escolar y el Uso de la computadora como herramienta de aprendizaje en alumnos del tercer y cuarto grado de Secundaria del Sector Privado*. USMP. Lima. Recuperado de: http://www.revistacultura.com.pe/imagenes/pdf/19_14.pdf
- López, M. (2013). *Aprendizaje, competencias y TIC, aprendizaje basado en competencias*. (Pearson, Ed.). México.
- Luaran, J. E., Binti, F., Mohd, K., y Nadzri, F. Luaran, J. E. (2011). Hooked on the internet: How does it influence the quality of undergraduate student's academic performance? 2 . 0 Review of Related Literature.
- Luna, R. (2012). *Incidencia del uso de tecnologías en las actividades académicas sobre el rendimiento académico*. Universidad Técnica Particular de Loja. Recuperado de: [http://dspace.utpl.edu.ec/bitstream/123456789/3113/1/Luna Vasquez Ruth Elizabeth.pdf](http://dspace.utpl.edu.ec/bitstream/123456789/3113/1/Luna%20Vasquez%20Ruth%20Elizabeth.pdf)
- Malhotra, N. . (1997). Investigación de Mercados un Enfoque práctico. México: Pearson Education, 2.
- Martínez, C. M., y Sepúlveda, M. A. R. (2012). Introducción al análisis factorial exploratorio. *Revista Colombiana de Psiquiatría*, 41(1), 197-207. Recuperado de: <http://www.redalyc.org/resumen.oa?id=80624093014>
- Meso, K., Pérez, J., y Mendiguren, T. (2011). Implementación de las Redes Sociales en la Enseñanza Superior Universitaria. *Tejuelo: Didáctica de la Lengua y la Literatura. Educación*, 12, 137-155.
- Molina, J., y García, J. (2006). *Técnicas de Análisis de datos: Aplicaciones Prácticas utilizando Microsoft excel y weka*. Universidad Carlos III, Madrid.España.
- Morales, P. (2013). *El Análisis Factorial en la construcción e interpretación de tests, escalas y cuestionarios*. Facultad de Ciencias Humanas y Sociales. Madrid.España. Recuperado de: <http://web.upcomillas.es/personal/peter/investigacion/AnalisisFactorial.pdf>
- Moreno, M., Miguel, L., García, F., y Polo, J. (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. *Potencia*, (21812), 1-16. Recuperado de: <http://ceur-ws.org/Vol-84/paper4.pdf>
- Muñoz, J. (2013). *Incidencia de la política pública enmarcada a las tecnologías de información y comunicación (TIC) para la disminución de brechas y analfabetismo digital*. Pontificia Universidad Católica del Ecuador.

- Muñoz, P., y González, M. (2010). Estudio cuantitativo sobre el uso docente de herramientas teleformativas en el ámbito de la programación y base de datos. *EDUTECH, Revista Electrónica de Tecnología Educativa*, 32. Recuperado de: http://edutech.rediris.es/Revelec2/revelec32/estudio_cuantitativo_usodocente_herramientas_teleformativas_programacion_basedatos.html
- Naso, F., Balbi, M., Di Gracia, N., y Peri, J. (2012). La importancia de las Redes sociales en el ámbito educativo. Buenos Aires. Recuperado de: http://sedici.unlp.edu.ar/bitstream/handle/10915/18296/Documento_completo___pdf?seque nce=1
- Ordóñez, K. (2013). *Aplicación de técnicas de minería de datos para predecir la diserción de los estudiantes de la modalidad abierta y a distancia*. Universidad Técnica Particular de Loja. Recuperado de: http://dspace.utpl.edu.ec/bitstream/123456789/7897/1/Ordonez_Briceño_Karla- Informatica.pdf
- Osorio, G. (2009). *Análisis de características del ambiente creativo en empresas de Manizales con técnicas KDD*. Universidad Nacional de Colombia.
- Osorio-Zuluaga, G. A. (2009). *Análisis de características del ambiente creativo en empresas de Manizales con técnicas KDD*. Universidad Nacional de Colombia.
- Pando Fernández, V., y San Martín Fernandez, R. (2004). Regresión logística multinomial. *Cuad. Soc. Esp. Cien. For.*, 000(18), 323-327.
- Peña, D. (2002). *Análisis de datos Multivariantes*. (McGraw-Hill Interamericana de España, Ed.). España: 23 de enero de 2002. Recuperado de: http://civil2011.bligoo.cl/media/users/15/791368/files/138227/An_lisis_de_Datos_Multivaria ntes_-_Daniel_Pe_a.pdf
- Pérez, C. (2009). *Técnicas estadísticas multivariantes con SPSS*. Madrid: *Garceta grupo editorial*.
- Pérez, C., y Santín, D. (2007). *Minería de datos: técnicas y herramientas*. (E. Paraninfo, Ed.) (reimpresa.).
- Pérez, E., y Medrano, L. (2010). Análisis Factorial Exploratorio: Bases Conceptuales y Metodológicas. *Revista Argentina de Ciencias del Comportamiento (RACC)*, 2(1), 58-66.
- Pérez García, Á. (2013). Redes Sociales y Educación Una reflexión acerca de su uso didáctico y creativo. *Creatividad y sociedad: revista de la Asociación para la Creatividad*, (21), 12-23. Recuperado de: <http://dialnet.unirioja.es/servlet/articulo?codigo=4786492yinfo=resumenyidioma=SPA>
- Pérez, M. (2014). *Minería de datos a traves de ejemplos*. (RC libros, Ed.). Madrid.España
- Poch, F. V. (2009). Uso autoinformado de Internet en adolescentes: Perfil psicológico de un uso elevado de la red. *International Journal of Psychology and Psychological Therapy*, 9(1), 109-122.

- Quintín, M., y Paz, Y. del R. (2007). *Tratamiento estadístico de datos con SPSS*. (Editorial Parainfo, Ed.). Madrid.España.
- Raines, J. M. (2012). The Effect of Online Homework Due Dates on College Student Achievement in Elementary Algebra. *Journal of Studies in Education*, 2(3), 1-18. doi:10.5296/jse.v2i3.1704
- Rencher, A. C. (2002). *Methods of Multivariate Analysis*. New York: WileyInterscience. (2a ed.).
- Ruiz, E., y Ortiz, H. (2014). Acceso a Internet e Impacto en los Hogares Peruanos . Una Evaluación a Partir de Mircrodatos. *Revista de estudios para el desarrollo social de la Comunicación*, (9), 361-375.
- Ryan, T. P. (1997). *Modern regression methods*. Wiley.
- Salazar, M. (2013). *Las redes sociales de internet y su incidencia en el rendimiento académico de los estudiantes del décimo año de educacion básica*. 14- Junio 2013. Colegio Menor Indoamérica. Recuperado de: http://repo.uta.edu.ec/bitstream/123456789/4250/1/tse_2012_43.pdf
- Salinas, J. (2013). Enseñanza Flexible y Aprendizaje Abierto, Fundamentos clave de los PLEs. *Entornos personales de aprendizaje: Claves para el ecosistema educativo en red*, 53-70. Recuperado de: <http://digitum.um.es/jspui/bitstream/10201/30410/1/capitulo3.pdf>
- Santos, M. (2010). Análisis de redes sociales y rendimiento académico:lecciones a partir del caso de los Estados Unidos. *Revista debates en sociología*, (35), 7-44. Recuperado de: <http://revistas.pucp.edu.pe/index.php/debatesensociologia/article/view/2125/2057>
- Sarasa, R. B., Suárez, A. R., y Sánchez, R. A. (2008). *Desarrollo de un proceso de KDD en el ámbito docente : Preparación de los datos*. La Habana. Cuba. Recuperado de: <http://ccia.cujae.edu.cu/index.php/siia/siia2008/paper/download/1139/231>.
- Serrano, A., y Martinez, E. (2003). *La Brecha Digital: Mitos y Realidades"*. (UABC, Ed.)*Editorial Universitaria de l Universidad Autónoma de Naja California*.
- Sierra-Araujo, B. (2006). *Aprendizaje Automático:conceptos basicos y avanzados*. (PRENTICE-HALL, Ed.). Madrid.España.
- Tamez, P. de J. (2012). *Adicción a la Red Social de Facebook y su incidencia en el Rendimiento*. Recuperado de: <http://cdigital.dgb.uanl.mx/te/1080256440.PDF>
- Taucher, E. (1997). *Bioestadística*. (Editorial Universitaria, Ed.)*Bioestadística*.
- Taylor, Zhu, Dekkers, y Marshall. (2003). Socio-Economic Factors Affecting Home Internet Usage Patterns in Central Queensland. *InformingScience Journal*, 6, 233-246.
- Tello, E., Sosa, Cl., Castillo, L., y Flores, M. (2015). Análisis de los servicios de la tecnología Web 2 . 0 aplicados a la educación, 1-12.

- 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.
- UNESCO. (2007). *Mediun-term strategy 2002-2007*. Recuperado de: <http://unesdoc.unesco.org/images/0012/001254/125434e.pdf>
- Unión Internacional de Telecomunicaciones. (2012). *Medición de la sociedad de la Información 2012*. Recuperado de: https://www.itu.int/dms_pub/itu-d/opb/ind/D-IND-ICTOI-2012-SUM-PDF-S.pdf
- Valenzuela, R. (2013). Las redes sociales y su aplicación en la educación. *Revista Digital Universitaria*, 14(4), 1-14. Recuperado de: <http://www.revista.unam.mx/vol.14/num4/art36/art36.pdf>
- Vallejo, D., y Tenelanda, G. (2012). Minería de datos aplicada en detención de intrusos. *Ing. USBMed*, 3(1), 50-61. Recuperado de: <http://web.usbmed.edu.co/usbmed/fing/v3n1/v3n1a6.pdf>
- Vallejo, S. (2006). *Minería de datos*. Corrientes - Argentina. Recuperado de: http://exa.unne.edu.ar/informatica/SO/Mineria_Datos_Vallejos.pdf
- Vázquez, A. (2014). *Incidencia de la brecha digital en grupos de iguales a partir de la interactividad entre la identidad física y la identidad digital*. (Lulu.com, Ed.). Macedonia.
- Vicente, J. (2011). *Introducción al Análisis de Cluster*. Recuperado de: <http://benjamindespensa.tripod.com/spss/AC.pdf>
- Vieira, L. P., Ortiz, L. I., y Ramirez, S. (2009). *Introducción a la Minería de datos*. (Editora E-papers, Ed.). Río de Janeiro- Brasil.
- Vilà, R., Rubio, M. J., Berlanga, V., y Torrado, M. (2014). Cómo aplicar un cluster jerárquico en SPSS. *REIRE. Revista d'Innovació i Recerca en Educació*, 7(1), 113-127. doi:10.1344
- Vizcaino, P. (2008). *Aplicación de técnicas de inducción de árboles de decisión a problemas de clasificación mediante el uso de WEKA*. Fundación Universitaria Konrad Lorenz. Recuperado de: http://www.konradlorenz.edu.co/images/stories/suma_digital_sistemas/2009_01/final_paula_andrea.pdf
- Warschauer, M. (2008). *Laptops and literacy: a multi-site case study*.
- Wittwer, J., y Senkbeil, M. (2005). Is students' computer use at home related to their mathematical performance at school? *Computers y Education*, 50, 1558–1571. Recuperado de: [http://www.pgce.soton.ac.uk/ict/SecondaryICT/PDFs/Student computer use at home and mathematics performance at school secondary.pdf](http://www.pgce.soton.ac.uk/ict/SecondaryICT/PDFs/Student%20computer%20use%20at%20home%20and%20mathematics%20performance%20at%20school%20secondary.pdf)
- Xu, R., y Wunsch, D. C. (2009). *Clustering*. New Jersey: Wiley-IEEE Press.

Yu, P. (2002). Equality in the information age: forward. *Cardozo and entertainment law journal.*, 20.
Recuperado de:
<http://scholarship.law.tamu.edu/cgi/viewcontent.cgi?article=1476&context=facscholar>

Zillien, N., y Hargittai, E. (2009). Digital Distinction: Status-Specific Types of Internet Usage. *Social Science Quarterly*, 90(2), 274-291. doi:10.1111/j.1540-6237.2009.00617.x

ANEXOS

Anexo 1: Población universitaria del periodo académico Octubre 2014- Febrero 2015

Total de estudiantes matriculados periodo Septiembre 2014 Febrer...

https://mail.google.com/mail/#inbox/14bdfc3518c0d805

Correo

REDACTAR

Recibidos
Destacados
Enviados
Borradores
MIOS
UNL
Más



¡Vaya! No eres invisible porque has accedido a Google Talk desde otro cliente, dispositivo o ubicación que no es compatible con la invisibilidad.

Buscar contactos...

flacaamgb@gmail.com quiere chatear contigo. ¿Aceptas?

Milton Ricardo Pal...
Ernesto Perez
José Lubín Sando...
jorge luis malla sa...
Milton Labanda Unl
Ruth Carmita Flor...
Sandra del Carme...
Diego Mora

Total de estudiantes matriculados periodo Septiembre 20

José Javier Martínez

para mí

Total de estudiantes: 8318

Saludos



Haz clic aquí si quieres [Responder](#) o [Reenviar](#) el mensaje

0,06 GB en uso
[Administrar](#)

©2015 Google - [Término](#)

Con la tec



Anexo 2: Encuesta para estudiantes

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									
¿Qué carrera estudia?									
3. Responda la siguiente pregunta									
¿Cuál es su edad?									
4. Responda la siguiente pregunta									
¿Cuál es su género?									
Hombre				Mujer					
()				()					
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									
De 1 a 7, ¿cuántos días a la semana se conecta Internet?									
1	2	3	4	5	6	7			
()	()	()	()	()	()	()	()		
8. Responda las siguientes preguntas									
De 1 a 10 su nivel de conocimientos en el manejo de Internet es:									
1	2	3	4	5	6	7	8	9	10
()	()	()	()	()	()	()	()	()	()
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 tienen 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 (iPod; 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 3: Carta dirigida al rector de la Universidad en estudio.

Loja, 5 de Junio del 2014

Sr. Dr.

Gustavo Villacís Rivas

RECTOR DE LA UNIVERSIDAD NACIONAL DE LOJA

Por medio de la presente, yo Tatiana Elizabeth Jaramillo Tapia con número de cédula 1104101314, egresada de la Universidad Técnica Particular de Loja, en la Modalidad Abierta y a Distancia, en la carrera de Informática.

Me pongo en contacto con usted para manifestarle mi deseo e interés de realizar el recogimiento y levantamiento de información (encuestas y entrevistas) para el desarrollo de mi tesis con el título "Educación Virtual en el Ecuador"; en cada una de las áreas que tiene esta prestigiosa Universidad.

El objetivo central de mi tesis se refiere a realizar una investigación sobre el uso de tecnología en los procesos académicos de una universidad, por tanto me he permitido escoger a la Universidad Nacional de Loja como base para el desarrollo de mi proyecto de tesis.

Por todo ello, solicito su autorización para desarrollar este estudio y me gustaría contar con su colaboración, así como con la del resto de la comunidad educativa, para el desarrollo de esta investigación, en lo que pudiese resultar necesario.

Agradezco desde ya la atención que brinde a mi carta y de ser aceptada les agradeceré me lo hagan saber a través de mi e-mail tatys0623@gmail.com, le saluda atentamente:

Tatiana E. Jaramillo Tapia

Egresada de Informática en la Modalidad Abierta y a Distancia

Universidad Técnica Particular de Loja

tatys0623@gmail.com

Anexo 4: Contestación favorable de la carta dirigida al rector de la Universidad en estudio.



UNIVERSIDAD NACIONAL DE LOJA
ACREDITADA-Resolución N°003-CONEA-2010-111-DC

Of. Nro. **020142369**-R-UNL
Loja, 14 de julio de 2014

Señores
DIRECTORES DE LAS AREAS UNIVERSITARIAS
Ciudad.-

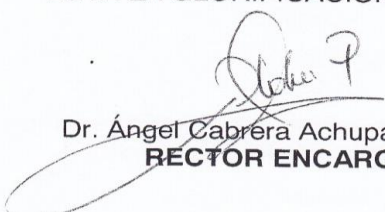
De mi consideración:

En relación al trámite Nro. 202015, de fecha 9 de junio de 2014, la señorita Tatiana Jaramillo Tapia, egresada de Información en la Modalidad Abierta y a Distancia de la Universidad Técnica Particular de Loja, solicita autorización para realizar recogimiento y levantamiento de información del tema educación virtual en el Ecuador para su tesis de grado en la UTPL.

Al respecto, me permito dirigir a usted con la finalidad de solicitarle comedidamente se atienda el pedido de la Srta. Jaramillo; y, se brinde la colaboración necesaria, para que realice encuestas y entrevistas en cada una de las áreas de nuestra Alma Mater, a partir del próximo período académico 2014 – 2015.

Con los sentimientos de mi especial consideración.

Muy atentamente,
EN LOS TESOROS DE LA SABIDURIA
ESTA LA GLORIFICACION DE LA VIDA


Dr. Ángel Cabrera Achupallas, Mg.Sc.
RECTOR ENCARGADO

cc. Secretaría del Rectorado,
Archivo.



Ciudad Universitaria "Guillermo Falconi Espinosa" Casilla letra "S"
Teléfonos: 2547-252-Ext.101; 2547-200
E-Mail: rector@unl.edu.ec Loja-Ecuador

Anexo 5: Análisis Factorial para las actividades académicas

Matriz de correlaciones	
a.	Determinante = ,375

KMO y prueba de Bartlett

Medida de adecuación muestral de Kaiser-Meyer-Olkin.		,762
Prueba de esfericidad de Chi-cuadrado aproximado		489,357
Bartlett	gl	36
	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	2,571	28,562	28,562	2,571	28,562	28,562	1,685	18,722
2	1,038	11,533	40,095	1,038	11,533	40,095	1,524	16,931	35,653
3	1,001	11,120	51,215	1,001	11,120	51,215	1,401	15,561	51,215
4	,923	10,254	61,469						
5	,827	9,190	70,659						
6	,774	8,598	79,256						
7	,670	7,447	86,704						
8	,654	7,265	93,969						
9	,543	6,031	100,000						

Método de extracción: Análisis de Componentes principales.

Anexo 6: Análisis Factorial para entretenimiento

Matriz de correlaciones
a. Determinante = ,439

KMO y prueba de Bartlett

Medida de adecuación muestral de Kaiser-Meyer-Olkin.	,597
Prueba de esfericidad de Chi-cuadrado aproximado	413,015
Bartlett	
gl	3
Sig.	,000

Varianza total explicada

Componente	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	1,953	65,112	65,112	1,953	65,112	65,112
2	,745	24,847	89,958			
3	,301	10,042	100,000			

Método de extracción: Análisis de Componentes principales.

Anexo 7: Análisis clúster para clasificar en dos grupos según el uso del Internet en actividades académicas.

Centros de los conglomerados finales

	Conglomerado	
	1	2
¿Aproximadamente cuántos recursos educativos descarga de la plataforma virtual cada mes?	3,71	5,40
¿Aproximadamente cuántas consultas les hace a sus profesores cada mes?	5,12	8,09
¿Aproximadamente cuántas consultas realiza a sus compañeros por mes?	6,70	12,67
¿Aproximadamente cuántos videos académicos mira en youtube cada mes?	6,30	9,00
¿Aproximadamente cuántas horas busca información académica en Internet cada mes?	10,33	36,10

ANOVA

	Conglomerado		Error		F	Sig.
	Media cuadrática	gl	Media cuadrática	gl		
¿Aproximadamente cuántos recursos educativos descarga de la plataforma virtual cada mes?	270,899	1	21,442	502	12,634	,000
¿Aproximadamente cuántas consultas les hace a sus profesores cada mes?	829,630	1	40,815	502	20,327	,000
¿Aproximadamente cuántas consultas realiza a sus compañeros por mes?	3360,130	1	50,838	502	66,094	,000
¿Aproximadamente cuántos videos académicos mira en youtube cada mes?	690,796	1	29,631	502	23,313	,000
¿Aproximadamente cuántas horas busca información académica en Internet cada mes?	62756,006	1	66,075	502	949,764	,000

Número de casos en cada conglomerado

Conglomerado	1	378,000
	2	126,000
Válidos		504,000
Perdidos		,000

Anexo 8: Análisis clúster para clasificar en tres grupos según el uso del Internet en actividades académicas.

Centros de los conglomerados finales

	Conglomerado		
	1	2	3
¿Aproximadamente cuántos recursos educativos descarga de la plataforma virtual cada mes?	4,76	6,40	3,00
¿Aproximadamente cuántas consultas les hace a sus profesores cada mes?	6,83	11,26	3,30
¿Aproximadamente cuántas consultas realiza a sus compañeros por mes?	9,20	15,79	4,68
¿Aproximadamente cuántos videos académicos mira en youtube cada mes?	8,68	9,34	5,49
¿Aproximadamente cuántas horas busca información académica en Internet cada mes?	41,27	19,50	8,82

ANOVA

	Conglomerado		Error		F	Sig.
	Media cuadrática	gl	Media cuadrática	gl		
¿Aproximadamente cuántos recursos educativos descarga de la plataforma virtual cada mes?	532,175	2	19,901	501	26,741	,000
¿Aproximadamente cuántas consultas les hace a sus profesores cada mes?	2846,309	2	31,190	501	91,257	,000
¿Aproximadamente cuántas consultas realiza a sus compañeros por mes?	5504,641	2	35,672	501	154,313	,000
¿Aproximadamente cuántos videos académicos mira en youtube cada mes?	798,780	2	27,880	501	28,650	,000
¿Aproximadamente cuántas horas busca información académica en Internet cada mes?	34436,047	2	53,999	501	637,711	,000

Número de casos en cada conglomerado

Conglomerado	1	82,000
	2	126,000
	3	296,000
Válidos		504,000
Perdidos		,000

Anexo 9: Análisis clúster para clasificar en cuatro grupos según el uso del Internet en actividades académicas.

Centros de los conglomerados finales

	Conglomerado			
	1	2	3	4
¿Aproximadamente cuántos recursos educativos descarga de la plataforma virtual cada mes?	6,93	5,02	2,53	4,93
¿Aproximadamente cuántas consultas les hace a sus profesores cada mes?	12,64	8,58	3,81	4,72
¿Aproximadamente cuántas consultas realiza a sus compañeros por mes?	20,92	11,71	4,91	5,66
¿Aproximadamente cuántos videos académicos mira en youtube cada mes?	9,71	9,77	5,07	7,64
¿Aproximadamente cuántas horas busca información académica en Internet cada mes?	20,03	48,92	6,12	21,36

ANOVA

	Conglomerado		Error		F	Sig.
	Media	gl	Media	gl		
	cuadrática		cuadrática			
¿Aproximadamente cuántos recursos educativos descarga de la plataforma virtual cada mes?	441,068	3	19,423	500	22,708	,000
¿Aproximadamente cuántas consultas les hace a sus profesores cada mes?	1672,880	3	32,600	500	51,315	,000
¿Aproximadamente cuántas consultas realiza a sus compañeros por mes?	5449,372	3	25,066	500	217,404	,000
¿Aproximadamente cuántos videos académicos mira en youtube cada mes?	615,816	3	27,436	500	22,445	,000
¿Aproximadamente cuántas horas busca información académica en Internet cada mes?	26611,861	3	32,181	500	826,956	,000

Número de casos en cada conglomerado

Conglomerado	1	76,000
	2	48,000
	3	232,000
	4	148,000
Válidos		504,000
Perdidos		,000

Anexo 10: Análisis clúster para clasificar en dos grupos según los usos de la tecnología para el entretenimiento.

Centros de los conglomerados finales

	Conglomerado	
	1	2
¿Aproximadamente cuántas horas a la semana chatea por diversión?	5,00	17,97
¿Aproximadamente cuántas horas a la semana utiliza redes sociales?	6,45	23,14
¿Aproximadamente cuántos videos para entretenimiento mira en youtube cada semana?	5,37	11,38

ANOVA

	Conglomerado		Error		F	Sig.
	Media cuadrática	gl	Media cuadrática	gl		
¿Aproximadamente cuántas horas a la semana chatea por diversión?	15554,687	1	23,153	502	671,818	,000
¿Aproximadamente cuántas horas a la semana utiliza redes sociales?	25754,764	1	29,014	502	887,657	,000
¿Aproximadamente cuántos videos para entretenimiento mira en youtube cada semana?	3340,582	1	37,560	502	88,939	,000

Número de casos en cada conglomerado

Conglomerado	1	382,000
	2	122,000
Válidos		504,000
Perdidos		,000

Anexo 11: Análisis clúster para clasificar en tres grupos según los usos de la tecnología para el entretenimiento.

Centros de los conglomerados finales

	Conglomerado		
	1	2	3
¿Aproximadamente cuántas horas a la semana chatea por diversión?	9,41	19,49	3,30
¿Aproximadamente cuántas horas a la semana utiliza redes sociales?	11,40	26,55	4,27
¿Aproximadamente cuántos videos para entretenimiento mira en youtube cada semana?	8,73	12,14	3,65

ANOVA

	Conglomerado		Error		F	Sig.
	Media cuadrática	gl	Media cuadrática	gl		
¿Aproximadamente cuántas horas a la semana chatea por diversión?	8564,247	2	20,058	501	426,974	,000
¿Aproximadamente cuántas horas a la semana utiliza redes sociales?	15932,426	2	16,876	501	944,063	,000
¿Aproximadamente cuántos videos para entretenimiento mira en youtube cada semana?	2771,810	2	33,238	501	83,392	,000

Número de casos en cada conglomerado

Conglomerado	1	171,000
	2	86,000
	3	247,000
Válidos		504,000
Perdidos		,000

Anexo 12: Análisis clúster para clasificar en cuatro grupos según los usos de la tecnología para el entretenimiento.

Centros de los conglomerados finales

	Conglomerado			
	1	2	3	4
¿Aproximadamente cuántas horas a la semana chatea por diversión?	7,74	23,05	15,50	4,29
¿Aproximadamente cuántas horas a la semana utiliza redes sociales?	11,87	19,34	31,19	5,27
¿Aproximadamente cuántos videos para entretenimiento mira en youtube cada semana?	12,73	9,67	11,10	3,44

ANOVA

	Conglomerado		Error		F	Sig.
	Media cuadrática	gl	Media cuadrática	gl		
¿Aproximadamente cuántas horas a la semana chatea por diversión?	6514,004	3	15,271	500	426,558	,000
¿Aproximadamente cuántas horas a la semana utiliza redes sociales?	10256,793	3	19,099	500	537,029	,000
¿Aproximadamente cuántos videos para entretenimiento mira en youtube cada semana?	2813,106	3	27,513	500	102,246	,000

Número de casos en cada conglomerado

Conglomerado	1	110,000
	2	58,000
	3	42,000
	4	294,000
Válidos		504,000
Perdidos		,000

Anexo 13: Análisis clúster para clasificar en dos grupos según el nivel de uso de los dispositivos

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
Cámara digital	10,00	1,00
Teléfono móvil sin acceso a Internet	10,00	1,00

Número de casos en cada conglomerado

Conglomerado	1	238,000
	2	266,000
Válidos		504,000
Perdidos		,000

Anexo 14: Análisis clúster para clasificar en tres grupos según el nivel de uso de los dispositivos.

Centros de los conglomerados finales

	Conglomerado		
	1	2	3
Smartphone con cámara fotográfica y acceso a Internet	2,09	2,92	7,93
Teléfono móvil con acceso a Internet	2,75	4,76	8,95
Teléfono móvil sin acceso a Internet	8,86	1,90	5,12
Computador portátil	7,34	5,89	8,78
Cámara digital	3,54	2,22	6,86

Número de casos en cada conglomerado

Conglomerado	1	110,000
	2	193,000
	3	201,000
Válidos		504,000
Perdidos		,000

Anexo 15: Análisis clúster para clasificar en cuatro grupos según el nivel de uso de los dispositivos

Centros de los conglomerados finales

	Conglomerado			
	1	2	3	4
Smartphone con cámara fotográfica y acceso a Internet	7,69	2,19	7,84	1,91
Teléfono móvil con acceso a Internet	8,76	4,13	8,64	2,76
Teléfono móvil sin acceso a Internet	2,04	1,90	9,01	8,87
Computador portátil	8,22	5,77	9,01	7,17
Cámara digital	4,79	2,30	8,53	3,29

Número de casos en cada conglomerado

Conglomerado	1	144,000
	2	162,000
	3	92,000
	4	106,000
Válidos		504,000
Perdidos		,000

Anexo 16: Ajuste del modelo del uso de Internet en actividades académicas.

Información del ajuste del modelo

Modelo	Criterio de ajuste del modelo			Contrastes de la razón de verosimilitud		
	AIC	BIC	-2 log verosimilitud	Chi-cuadrado	gl	Sig.
Sólo la intersección	69,632	78,077	65,632			
Final	62,520	104,746	42,520	23,112	8	,003

Bondad de ajuste

	Chi-cuadrado	gl	Sig.
Pearson	,000	0	.
Desviación	,000	0	.

Pseudo R-cuadrado

Cox y Snell	,037
Nagelkerke	,044
McFadden	,020

Anexo 17: Ajuste del modelo en función de los usos de tecnología en el entretenimiento

Información del ajuste del modelo

Modelo	Criterio de ajuste del modelo			Contrastes de la razón de verosimilitud		
	AIC	BIC	-2 log verosimilitud	Chi-cuadrado	gl	Sig.
Sólo la intersección	58,600	67,045	54,600			
Final	63,638	105,863	43,638	10,962	8	,204

Bondad de ajuste

	Chi-cuadrado	gl	Sig.
Pearson	,000	0	.
Desviación	,000	0	.

Pseudo R-cuadrado

Cox y Snell	,022
Nagelkerke	,025
McFadden	,011

**Anexo 18: Ajuste del modelo del uso de la tecnología en el aprendizaje sobre el
rendimiento académico**

Pruebas omnibus sobre los coeficientes del modelo

		Chi cuadrado	gl	Sig.
Paso 1	Paso	3,988	1	,046
	Bloque	3,988	1	,046
	Modelo	3,988	1	,046

Prueba de Hosmer y Lemeshow

Paso	Chi cuadrado	gl	Sig.
1	2,666	1	,103

Tabla de clasificación^a

Observado		Pronosticado		
		En el semestre anterior, aprobo o no aprobo todas las asignaturas?		Porcentaje correcto
		Aprobo	No Aprobo	
Paso 1	En el semestre anterior, aprobo o no aprobo todas las asignaturas?	Aprobo	No Aprobo	
	Aprobo	485	0	100,0
	No Aprobo	19	0	,0
	Porcentaje global			96,2

a. El valor de corte es .500

Anexo 19: Ajuste del modelo del uso de la tecnología para entretenimiento sobre el rendimiento académico

Pruebas omnibus sobre los coeficientes del modelo

		Chi cuadrado	gl	Sig.
Paso 1	Paso	,682	1	,409
	Bloque	,682	1	,409
	Modelo	,682	1	,409

Prueba de Hosmer y Lemeshow

Paso	Chi cuadrado	gl	Sig.
1	,003	1	,956

Tabla de clasificación^a

Observado		Pronosticado		
		En el semestre anterior, aprobo o no aprobo todas las asignaturas?		Porcentaje correcto
		Aprobo	No Aprobo	
Paso 1	En el semestre anterior, aprobo o no aprobo todas las asignaturas?	Aprobo	No Aprobo	Porcentaje global
	Aprobo	485	0	100,0
	No Aprobo	19	0	,0
	Porcentaje global			96,2

a. El valor de corte es .500

Anexo 20: Información General del proyecto “Educación virtual en Ecuador”

Universidad Técnica Particular de Loja

TITULACIÓN DE INGENIERÍA EN SISTEMAS INFORMÁTICOS Y COMPUTACIÓN

Guía para elaboración de Propuestas de Proyectos de Fin de Titulación

Información General del Proyecto

Título del proyecto:	Educación virtual en Ecuador
Duración:	9 meses
Propuesto por:	Equipo: Juan Carlos Torres Díaz Carlos Correa Granda Héctor Gómez Inés Jara Diego Alvarado Daisy Karina García Lorena Condolo Martha Agila Pablo Torres Rosario Requena Paola Andrade
Docente Investigador:	Juan Carlos Torres Díaz (jctorres@utpl.edu.ec)
Línea de Investigación:	Tecnología aplicada a la educación

Perfil Requerido del estudiante	35 Egresados de Ingeniería en Informática
---------------------------------	---

Propósito / Descripción

El proyecto busca levantar información de las universidades ecuatorianas de las categorías A, B y C, esta información permitirá contar con una línea de base respecto al uso de la tecnología en el sistema educativo del país, esta información es fundamental para la toma de decisiones que permitan optimizar el uso de la tecnología y maximizar sus efectos en el rendimiento académico de los estudiantes.

Se levantará información referente a como los estudiantes utilizan la tecnología en sus actividades académicas y de entretenimiento. Se buscará patrones de uso de la tecnología y se categorizará a los estudiantes en grupos homogéneos; así también se buscará la existencia de relaciones entre los usos de la tecnología y el éxito académico que alcanzan los estudiantes lo que permitirá determinar el real efecto de las herramientas informáticas dentro del sistema educativo.

Un tema sensible que requiere de un análisis sustentado es el que tiene que ver con el uso ético de la tecnología, esto incluye el plagio de información en la red y las implicaciones que tendría sobre los niveles de aprendizaje y el éxito académico.

Competencias a desarrollar

Levantar, tabular y organizar datos para la toma de decisiones
Inferir y generar información a partir de datos organizados

Componentes:

Componente	Descripción
Diseño de la investigación	En este componente se van a definir todos los aspectos metodológicos que se requieren previo inicio del trabajo, se definen aquí instituciones, tamaños de muestras, cuestionarios, procedimientos estadísticos de análisis. Este trabajo está a cargo del proponente del proyecto y será dado a conocer a los participantes.
Inducción a tesistas	Es una etapa de formación en la que se da a conocer de forma detallada la metodología a emplear, el procedimiento, aspectos legales, plazos y entregables.
Levantamiento y organización de datos	En este componente, los participantes encuestan a los estudiantes y profesores de las universidades, digitan los datos y realizan un análisis preliminar de los mismos.
Tratamiento y análisis de datos	En esta etapa se aplican procedimientos estadísticos a fin de obtener, entre otros resultados, categorizaciones, correlaciones, modelos de regresión. Con esos insumos se procede a documentar las relaciones y hallazgos.

Generación de informe	Es la etapa final, consiste en documentar el trabajo y consta de los siguientes apartados: Introducción Metodología Resultados Discusión de resultados Conclusiones Referencias
-----------------------	---

Estrategia o Metodología de desarrollo (Opcional)

Como parte inicial, la UTPL a través del director del proyecto diseñará el proyecto en sus aspectos metodológicos y organizativos. Posteriormente se procederá a instruir a los estudiantes y a poner en práctica los siguientes aspectos:

Aspecto 1: Autorización y levantamiento de datos para una muestra representativa en cada institución seleccionada

Aspecto 2: Utilización del cuestionario “usos de tecnología en las universidades” basado en los utilizados en los proyectos PIC, DLINHE, ECUADOR

Aspecto 3: Aplicación de análisis multivariante para obtener categorizaciones respecto a los usos de Internet para la educación y para el entretenimiento.

Aspecto 4: Aplicación de regresiones (lineales, logísticas) para obtener modelos explicativos y/o predictivos que expliquen las distintas realidades.

Aspecto 5: Levantamiento de información cualitativa utilizando las técnicas: entrevista semi-estructurada, cuestionarios y búsqueda de información institucional en la red.

Resultados esperados

Informe de investigación con los apartados siguientes:

Introducción: que se trata el trabajo / objetivos / preguntas de investigación / importancia del trabajo / componentes del trabajo

Marco teórico: revisión teórica de por lo menos 25 autores que traten los temas de brecha digital/usos de Internet/rendimiento académico/

Metodología: provista por la UTPL

Resultados: Descripción de los hallazgos

Discusión de los resultados:

Conclusiones.

Cronograma

En función de cada componente, indicar la duración estimada.

Componente	Tiempo
Diseño de la investigación	1 mes
Inducción y formación de los investigadores	1 mes
Levantamiento y organización de datos	2 meses
Tratamiento y análisis de datos	1 meses
Generación de informe	2 mes
Presentación y trámites legales	1 mes

Bibliografía

Torres-Díaz, J. C., Morocho, M., y Guamán, J. (2010). La educación virtual en Ecuador. In C. Rama y J. L. Pardo (Eds.), *La educación superior a distancia: Miradas diversas desde Iberoamérica* (pp. 76–91). Madrid: INTEVED.

Torres-Díaz, J. C., y Infante-Moro, A. (2011). Desigualdad digital en la universidad: Incidencia de los ingresos en los usos de Internet en Ecuador. *Comunicar*, 37(2). doi:10.3916/C37-2011-02-08

Anexo 1: Listado de universidades de categorías A, B y C

Universidades Categoría A

1. Escuela Politécnica Nacional
2. Escuela Superior Politécnica del Litoral
3. Universidad San Francisco de Quito

Universidades Categoría B

4. Escuela Superior Politécnica de Chimborazo
5. Pontificia Universidad Católica del Ecuador
6. Universidad Casa Grande Universidad Católica
7. Santiago de Guayaquil
8. Universidad Central del Ecuador
9. Universidad de Cuenca
10. Universidad del Azuay
11. Universidad Estatal de Milagro
12. Universidad Nacional de Loja
13. Universidad Particular Internacional Sek
14. Universidad Politécnica Salesiana
15. Universidad Técnica de Ambato
16. Universidad Técnica del Norte
17. Universidad Técnica Estatal de Quevedo
18. Universidad Técnica Particular de Loja
19. Universidad Tecnológica empresarial de Guayaquil
20. Universidad Tecnológica Equinoccial
21. Universidad Tecnológica Indoamérica

Universidades categoría C

22. Escuela Superior Politécnica agropecuaria de Manabí
23. Universidad de Especialidades Turísticas
24. Universidad de las Américas
25. Universidad del Pacífico – Escuela de Negocios
26. Universidad Estatal del Bolívar
27. Universidad Internacional del Ecuador
28. Universidad Laica Vicente Rocafuerte del Ecuador
29. Universidad Metropolitana
30. Universidad nacional de Chimborazo
31. Universidades de Especialidades Espíritu Santo
32. Universidad Regional Autónoma de los Andes
33. Universidad Técnica de Babahoyo
34. Universidad Técnica de Cotopaxi
35. Universidad Técnica de Israel