



UNIVERSIDAD TÉCNICA PARTICULAR DE LOJA

La Universidad Católica de Loja

ÁREA TÉCNICA

TÍTULO DE INGENIERO EN INFORMÁTICA

**Análisis de los usos de la tecnología en los estudiantes de la Universidad
Técnica del Norte**

TRABAJO DE TITULACIÓN

AUTORA: Jiménez Abad, Lady Susana

DIRECTORA: Santos Delgado, Ana Alexandra, Mgs.

**CENTRO UNIVERSITARIO CARIAMANGA
2016**

APROBACIÓN DEL DIRECTOR DEL TRABAJO DE TITULACIÓN

Magister.

Ana Alexandra Santos Delgado.

DOCENTE DE LA TITULACIÓN

De mi consideración:

El presente trabajo de titulación: Análisis de los usos de la tecnología en los estudiantes de la Universidad Técnica del Norte, realizado por Lady Susana Jiménez Abad, ha sido orientado y revisado durante su ejecución, por cuanto se aprueba la presentación del mismo.

Loja, Mayo 2016

f.....

DECLARACIÓN DE AUTORÍA Y CESIÓN DE DERECHOS

“Yo, Lady Susana Jiménez Abad declaro ser autora del presente trabajo de titulación: Análisis de los usos de la tecnología en los estudiantes de la Universidad Técnica del Norte, de la Titulación de Ingeniero en Informática, siendo la Mgs. Ana Alexandra Santos Delgado 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: “Formar parte del patrimonio de la Universidad la propiedad intelectual de investigaciones, trabajos científicos o técnicos y tesis de grado o trabajos de titulación que se realicen con el apoyo financiero, académico o institucional (operativo) de la Universidad”.

f.....
Lady Susana Jiménez Abad
C.I.: 1104542699

DEDICATORIA

Esta tesis quiero dedicarla primero a Dios y a la Virgen en su advocación a la Virgen del Cisne por siempre ser mi guía espiritual y mi confort cuando estoy a punto de desvanecer, por haberme permitido llegar hasta este punto y haberme dado salud para lograr mis objetivos, además de su infinita bondad y amor .

A mis padres Lizardo y Alidia, gracias a ellos soy lo que soy, por su apoyo, consejos, comprensión, amor y ayuda en los momentos más difíciles. Me han dado todo lo que soy como persona, mis valores, mis principios, mi carácter, mi empeño, mi perseverancia y mi coraje para conseguir mis objetivos, a ustedes gracias infinitas.

A mis hermanos Elizabeth, Ana, Nelson, Elvis, Ronald, por estar siempre presentes, acompañándome para poderme realizar personal y profesionalmente. A mis abuelitos Manuel, Mariana (†), Alipio (†) y Regina millón gracias por el amor, consejos y cuidados, por ser siempre un ejemplo de fortaleza y perseverancia.

A Carlos mi amor, amigo, compañero, gracias por siempre estar a mi lado ayudándome, motivándome para terminar con esta meta de mi vida, sin tu apoyo esto no hubiese sido lo mismo.

AGRADECIMIENTO

Quiero iniciar agradeciendo de manera muy especial a mis padres por todo su apoyo moral y económico, para terminar con esta etapa de mi vida, mi formación como profesional, sin su apoyo y sacrificio nada hubiese sido posible a ustedes mi ejemplo de que con perseverancia y fe todo se logra en la vida, mi más imperecedera gratitud.

De manera especial quiero agradecer a mi Directora de Tesis Magister Ana Alexandra Santos Delgado, que con sus directrices, conocimientos, consejos, esfuerzo y paciencia supo guiarme para una final culminación de este proyecto de investigación.

Al Dr. Juan Carlos Torres Díaz, por darme la oportunidad de ser parte de este proyecto de investigación, por cada uno de sus acertados conocimientos impartidos sin ningún tipo de egoísmo, por su paciencia, dedicación y entrega para la culminación del mismo, gracias.

A la Secretaria de la Carrera de Informática Modalidad a Distancia Lic. Lidia Villacís, por su apoyo presto siempre ayudarme en la solución de cada uno de los tramites durante toda mi etapa como estudiante de la UTPL, gracias por ser esa voz de aliento de continuar siempre adelante.

A mi amiga y compañera Raquel, gracias infinitas por ser parte de esta aventura desde el inicio por tu apoyo y motivación, por las alegrías, tristezas y desesperaciones por cada una de nuestras vivencias juntas.

Al personal docente de la UTPL, Modalidad a Distancia que fue parte de mi formación como profesional, gracias por las enseñanzas impartidas.

A todos ustedes gracias infinitas.

ÍNDICE DE CONTENIDOS

CARATULA	i
APROBACIÓN DEL DIRECTOR DEL TRABAJO DE TITULACIÓN.....	ii
DECLARACIÓN DE AUTORÍA Y CESIÓN DE DERECHOS.....	iii
DEDICATORIA	iv
AGRADECIMIENTO	v
ÍNDICE DE CONTENIDOS	vi
ÍNDICE DE FIGURAS.....	x
ÍNDICE DE TABLAS	xi
RESUMEN.....	1
ABSTRACT	2
CAPITULO I: INTRODUCCIÓN	3
1.1. Introducción.....	4
1.2. Objetivos	5
1.2.1. General:.....	5
1.2.2. Específicos:	5
1.3. Preguntas de investigación.....	5
1.4. Hipótesis.....	6
1.5. Organización del documento	6
CAPITULO II: MARCO TEÓRICO Y CONCEPTUAL	7
2.1. Apreciación de las Tecnologías de la Información y Comunicación.....	8
2.2. Introducción de brecha digital.....	8
2.3. Factores que generan la brecha digital.....	9
2.4. Tipos de brecha digital.....	12
2.4.1. Brecha digital de conectividad	12
2.4.2. Brecha digital de uso	13

2.4.3. Brecha digital conceptual.....	14
2.5. Brecha digital en Latinoamérica.....	14
2.6. Brecha digital en Ecuador.....	16
2.7. Redes sociales en la educación.....	17
2.8. Internet y la educación superior.....	18
2.9. Minería de datos.....	18
2.9.1. Tipos de datos.....	19
2.9.2. Tipos de modelos.....	21
2.9.3. La minería de datos y el proceso de descubrimiento de conocimientos en bases de datos.....	21
2.9.4. Reconocimientos de patrones.....	25
2.9.5. Clustering.....	26
CAPITULO III: METODOLOGÍA.....	27
3.1. Descripción de la población y tamaño de la muestra.....	28
3.2. Instrumentos de recolección de información.....	29
3.3. Métodos aplicados para el análisis e interpretación de datos.....	30
3.3.1. Fase de integración y recopilación de datos.....	30
3.3.2. Fase de selección, limpieza y transformación.....	32
3.3.3. Fase de Minería de datos.....	34
3.3.4. Fase de evaluación e interpretación.....	41
CAPITULO IV: ANÁLISIS DE RESULTADOS.....	43
4.1. Particularidades de la población universitaria.....	44
4.1.1. Descripción de la muestra.....	44
4.1.2. Preferencias de los estudiantes sobre uso del internet.....	47
4.2. Usos del internet en lo académico y entretenimiento.....	51
4.2.1. Usos de índole académico.....	51
4.2.2. Usos índole no académico o entretenimiento.....	58

4.2.3. Particularidades rendimiento académicos de los estudiantes.....	62
4.3. Perfiles de los estudiantes	63
4.3.1. Reducción de variables: Factorización.....	63
4.3.2. Clusterización	64
4.3.3. Representación del perfil académico	64
4.3.4. Representación del perfil de entretenimiento	67
4.4. Fase de evaluación e interpretación de datos.....	70
4.4.1. El nivel de ingresos determina como se utiliza internet para el aprendizaje	71
4.4.2. El nivel de ingresos determina como se utiliza internet para entretenimiento..	72
4.4.3. El uso de la tecnología en el aprendizaje incide en el rendimiento académico	74
4.4.4. El uso de la tecnología para entretenimiento incide en el rendimiento académico	76
4.5. Análisis y discusión de los resultados.....	77
4.5.1. Determinante del nivel de ingresos en el uso de internet en actividades académicas	77
4.5.2. Determinante del nivel de ingresos en el uso de internet en actividades entretenimiento.....	78
4.5.3. Incidencia del uso de internet en actividades académicas sobre rendimiento académico	79
4.5.4. Incidencia del uso de internet en actividades de entretenimiento sobre rendimiento académico.....	79
CONCLUSIONES Y RECOMENDACIONES	80
Conclusiones	81
Recomendaciones	82
BIBLIOGRAFÍA.....	83
ANEXOS.....	87
Anexo 1: Encuesta aplicada al estudiante	88
Anexo 2: Guía de entrevista para un(a) directivo de la institución educativa	91

Anexo 3: Carta para el levantamiento de información	92
Anexo 4: Carta de aceptación a levantamiento de información	93
Anexo 5: Análisis factorial para usos de internet en actividades académicas	94
Anexo 6: Análisis factorial usos de internet en actividades de entretenimiento.....	95
Anexo 7: Análisis factorial nivel de uso de dispositivos	96
Anexo 8: Uso de internet en actividades académicas clúster de dos grupos	97
Anexo 9: Uso de internet en actividades académicas clúster de tres grupos.....	98
Anexo 10: Uso de internet en actividades académicas clúster de cuatro grupos.....	99
Anexo 11: Uso de internet en actividades académicas clúster de cinco grupos	100
Anexo 12: Uso de internet en actividades de entretenimiento clúster de dos grupos.....	101
Anexo 13: Uso de internet en actividades de entretenimiento clúster de tres grupos	102
Anexo 14: Uso de internet en actividades de entretenimiento clúster de cuatro grupos	103
Anexo 15: Uso de internet en actividades de entretenimiento clúster de cinco grupos ..	104
Anexo 16: Análisis discriminante para la verificación del porcentaje de exactitud en actividades académicas	105
Anexo 17: Análisis discriminante para la verificación del porcentaje de exactitud en actividades de entretenimiento.....	107
Anexo 18: Modelo de Regresión Logística Binaria aplicada al perfil de uso de internet en actividades académicas	109
Anexo 19: Modelo de Regresión Logística Multinomial aplicada al perfil de uso de internet en actividades de entretenimiento	111
Anexo 20: Modelo de Regresión Logística Binaria aplicada al uso de internet en actividades académicas sobre el rendimiento académico	112
Anexo 21: Modelo de Regresión Logística Binaria aplicada al uso de internet en actividades de entretenimiento sobre el rendimiento académico	113
Anexo 22: Objetivos proporcionados por el director del proyecto	114

ÍNDICE DE FIGURAS

Figura 1: Proceso de KDD	22
Figura 2: Representación de las Tareas de Minería de Datos.....	23
Figura 3: Proceso del Reconocimiento de Patrones.....	26
Figura 4: Distribución de los estudiantes por género	44
Figura 5: Distribución de los estudiantes por titulación	45
Figura 6: Distribución estudiantil por edad	46
Figura 7: Distribución de los estudiantes por ingresos económicos familiares	46
Figura 8: Lugar habitual de conexión a internet	47
Figura 9: Días de conexión a internet	48
Figura 10: Ingreso a la plataforma virtual	52
Figura 11: Foros en los que participan mensualmente	55
Figura 12: Post o Tweets académicos	55
Figura 13: Horas que chatea sobre temas académicos.....	56
Figura 14: Niveles de acceso de los estudiantes a la biblioteca virtual.....	57
Figura 15: Horas empleadas a descargas de música, videos y programas	60
Figura 16: Amigos y seguidores de los estudiantes	62
Figura 17: Rendimiento académico.....	63
Figura 18: Centroides del perfil académico.....	66
Figura 19: Centroides de perfil de entretenimiento.....	69

ÍNDICE DE TABLAS

Tabla 1: Población de América Latina que cuenta con acceso a internet desde su hogar (en %) en el 2000	15
Tabla 2: Variables determinadas transformación de los datos	33
Tabla 3: Variables que intervienen en la comprobación de hipótesis	34
Tabla 4: Discriminante de 2 grupos para actividades académicas	40
Tabla 5: Discriminante de 3 grupos para actividades de entretenimiento.....	40
Tabla 6: Pruebas de Chi Cuadrado para la Hipótesis 2.....	42
Tabla 7: Pruebas de Chi Cuadrado para la Hipótesis 1, Hipótesis 3 e Hipótesis 4.....	42
Tabla 8: Conocimiento en Internet	49
Tabla 9: Horas diarias de conexión a internet	49
Tabla 10: Años de experiencia en internet.....	50
Tabla 11: Niveles de uso de dispositivos móviles	51
Tabla 12: Consultas a sus profesores.....	53
Tabla 13: Consultas que realiza a sus compañeros.....	53
Tabla 14: Descargas de recursos educativos	54
Tabla 15: Videos educativos que mira en YouTube cada mes.....	54
Tabla 16: Información académica que busca en internet	57
Tabla 17: Horas semanales que chatea por diversión.....	58
Tabla 18: Horas semanales que utiliza redes sociales.....	59
Tabla 19: Horas semanales para juegos en línea	60
Tabla 20: Videos de entretenimiento que mira en YouTube.....	61
Tabla 21. Nivel de exactitud para grupos de actividades académicas.....	65
Tabla 22: Discriminación de variables académicas.....	65
Tabla 23. Nivel de exactitud para grupos de entretenimiento.....	68
Tabla 24: Discriminación de variables de entretenimiento	68
Tabla 25: Coeficientes del modelo de regresión logística de los usos de internet en actividades académicas e ingresos.....	72
Tabla 26: Coeficientes del modelo de regresión logística de los usos de internet en actividades de entretenimiento e ingresos	74
Tabla 27: Coeficientes del modelo de regresión logística binaria (H3)	75
Tabla 28: Coeficientes del modelo de regresión logística binaria (H4)	76

RESUMEN

El proyecto de investigación: Análisis de los usos de la tecnología en los estudiantes de la Universidad Técnica del Norte, para alcanzar los objetivos planteados se lo desarrollo bajo el proceso de KDD, empleando técnicas de minería de datos como: análisis clúster y discriminante para la obtención de grupos homogéneos de los estudiantes con características diferenciadas en el uso de internet, y por último se aplicó regresión logística multinomial y binomial para obtener modelos para la comprobación de hipótesis, estas técnicas se aplicaron a datos que particularizan características de brecha digital; concretamente se analizó la incidencia de la variable, nivel de ingresos económicos familiares de los estudiantes sobre los usos de internet en actividades académicas y de entretenimiento, de los resultados obtenidos se determinó el nivel de relación que existe entre la variable uso de internet y rendimiento académico del estudiante.

PALABRAS CLAVES: brecha digital, uso de internet, incidencia, aprendizaje, ingresos económicos, rendimiento académico, proceso de KDD.

ABSTRACT

The research project: Analysis of the uses of technology in students from the Technical University of the North, to achieve the objectives it development under the KDD process using mining techniques data such as cluster analysis and discriminant for obtaining homogeneous groups of students with different characteristics in the use of internet, and last binomial logistic regression multinomial and applied to obtain models for hypothesis testing, these techniques were applied to data that particularize features digital divide; specifically the impact of the variable level of family income of students on the uses of Internet in academic activities and entertainment, the results obtained are analyzed the level of relationship between the variable Internet use and academic performance was determined of student.

KEYWORDS: digital divide, internet use, advocacy, learning, income, academic achievement, KDD process.

CAPITULO I: INTRODUCCIÓN

1.1. Introducción

La tecnología de la información y comunicación (TIC) han desempeñado y continúan desempeñando un papel fundamental en la configuración de nuestra sociedad y cultura. Su utilización ha provocado modificaciones en todos los campos del saber a ritmo vertiginoso, proporcionando una serie de herramientas y contextos de comunicación y de aprendizaje, de enorme potencialidad y ha obligado a redefinir incluso el concepto de la realidad, a partir de la posibilidad de construir nuevas realidades.(Gallardo, 2012).

En el mundo de la educación, y en especial en el de la educación superior, se ha avanzado con celeridad en la facilitación del acceso universal y el uso de las TIC, tanto a nivel institucional como en las relaciones comunicativas entre profesores y estudiantes. A pesar de ello, el impacto real del uso específico de Internet en el proceso de enseñanza y aprendizaje está siendo más laborioso y más complejo debido a las diferencias de percepción, no únicamente generacionales sino también educativas, entre estudiantes y profesores. (Duart, 2010).

Hemos visto diversos avances en lo que respecta a tecnología y en especial internet, y como este influye en la educación superior, pero en el Ecuador y de manera particular en la Universidad Técnica del Norte aún no se encuentra datos o indicadores confiables que demuestre el uso de internet en actividades tales como; entretenimiento y rendimiento académico, y como los ingresos económicos determinan el acceso y el uso de internet en la población estudiantil de nivel superior. Es por eso que el principal propósito de esta investigación, es obtener información que nos permita conocer que uso le están dando los estudiantes de las universidades ecuatorianas al internet.

En este trabajo se agrupa a los estudiantes en base a sus características o variables, las mismas que hacen referencia a años de experiencia como internauta, participaciones en *foros, post, tweets*, consultas a compañeros y profesores, descargas de videos y material académicos, ingresos a la plataforma virtual de su universidad, utilización de la biblioteca virtual, haciendo referencia a lo académico, haciendo agrupaciones de dos, denominados regulares y dedicados. La segunda agrupación se la realizo en base al uso de internet para el entretenimiento, con variables como: chat por diversión, redes sociales, juegos en línea, descargas de música, videos y programas, mirar videos en *youtube*, obteniendo como resultado tres grupos denominados intensivos, pasivos y activos. Relacionando todos estos resultados con los ingresos familiares que tienen los estudiantes para determinar cuál es su

rendimiento académico. Para la verificación de las hipótesis planteadas se realiza el análisis desde la perspectiva de minería de datos, específicamente bajo el proceso KDD (Descubrimiento del conocimiento en grandes volúmenes de datos).

1.2. Objetivos

Los objetivos planteados para el desarrollo de esta investigación son los siguientes:

1.2.1. General:

Analizar los usos de la tecnología en los estudiantes de la Universidad Técnica del Norte.

1.2.2. Específicos:

- Determinar la incidencia de los ingresos familiares en el rendimiento académico y entretenimiento.
- Identificar la incidencia del uso de tecnología en actividades académicas y de entretenimiento.
- Relacionar los usos de internet con el rendimiento académico y con los niveles de ingreso.

Los objetivos planteados por el director del proyecto están descritos en el (Anexo 22)

1.3. Preguntas de investigación

Las preguntas planteadas para esta investigación se describen a continuación:

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 la presente investigación:

Relacionadas a pregunta 1

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

Relacionadas a pregunta 2

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. Organización del documento

El documento está dividido de la siguiente manera, constando de cuatro capítulos cada uno con su respectivo contenido.

Capítulo I consiste en la descripción de la investigación, presentación de objetivos, alcance del proyecto, preguntas de investigación e hipótesis de las mismas.

Capítulo II, marco teórico, haciendo una revisión de la bibliografía referente a la investigación, como brecha digital usos de internet y efectos de la tecnología sobre el rendimiento académico y minería de datos.

Capítulo III, metodología, este capítulo abarca cada uno del proceso realizados para obtener los resultados y análisis como: cálculo de la muestra, análisis clúster, análisis de varianza, análisis de correlaciones, análisis discriminante y regresión logística, basados en el proceso de minería de datos, específicamente el proceso de KDD.

Capítulo IV, resultados de la formación e interpretación grupos de actividades académicas y entretenimiento, formados en base a los procesos descritos en el capítulo de metodología. En este capítulo se presenta un apartado resultados analizados y discusión de cada una de las hipótesis y los niveles de determinantes que tienen los ingresos en el uso de internet en lo académico como entretenimiento e incidencias que tiene la tecnología con el rendimiento académico.

CAPITULO II: MARCO TEÓRICO Y CONCEPTUAL

2.1. Apreciación de las Tecnologías de la Información y Comunicación

La tecnología de la información y comunicación (TIC) desde su aparición han desempeñado un papel fundamental en la transformación de nuestra sociedad y cultura. Su utilización ha provocado modificaciones en todos los campos del saber a ritmo vertiginoso, proporcionando una serie de herramientas y contextos de comunicación y de aprendizaje, de enorme potencialidad y ha obligado a redefinir incluso el concepto de la realidad, a partir de la posibilidad de construir nuevas realidades, (Gallardo, 2012).

En el mundo de la educación, y en especial en el de la educación superior, se ha avanzado con rapidez en la facilitación del acceso universal y el uso de las TIC, tanto a nivel institucional como en las relaciones comunicativas entre profesores y estudiantes. A pesar de ello, el impacto real del uso específico de Internet en el proceso de enseñanza y aprendizaje está siendo más laborioso y más complejo debido a las diferencias de percepción, no únicamente generacionales sino también educativas, entre estudiantes y profesores, (Duart, 2010).

Dentro de la desigualdad digital entre alumnos universitarios (Castaño Muñoz, 2010), sugiere que se debería analizar diferentes dimensiones tales como el acceso, la intensidad de uso, y las finalidades de uso que hacen este tipo de usuarios de internet, permitiendo analizar las determinantes de cada una de estas dimensiones y su impacto social.

2.2. Introducción de brecha digital

La brecha digital se define como la separación que existe entre las personas (comunidades, estados, países...) que utilizan las Tecnologías de Información y Comunicación (TIC) como una parte rutinaria de su vida diaria y aquellas que no tienen acceso a las mismas y que aunque las tengan no saben cómo utilizarlas según (Serrano & Martínez, 2003).

Según el autor (Tello, 2008), la brecha digital es sin duda uno de los primeros conceptos que permite hacer reflexión en torno al tema de impacto social que han tenido las tecnologías de la información y comunicación (TIC). De esta forma (DiMaggio & Harggitai, 2001) afirman que es posible analizarla la brecha digital como una visión dicotómica que hace distinción entre las personas que tienen acceso a internet y las que no, sin duda el uso de internet y las (TIC), ha permitido un sinnúmero de innovaciones y ciertos tipos de cambio en diferentes

ámbitos así como restricciones. (DiMaggio, Hargittai, Celeste, & Shafer, 2004), hacen referencia a los usos diferenciados que hay en el acceso a internet y el uso que le dan al mismo desde la perspectiva de diferentes factores tanto socioeconómicos y educación. Pero no fue antes de la investigación realizada por la Administración Nacional de las Telecomunicaciones e información (NTIA) de los Estados Unidos en los años de 1995-1999, donde por primera vez se utilizó el término “brecha digital” esta investigación consistía establecer diferencias de quienes tienen y no tienen acceso a internet, esto se dio debido a que a solo se favorecía a extractos socio-económicos más altos (DiMaggio, Hargittai, Neuman, & Robinson, 2001), estas diferencias consistían básicamente en la disponibilidad de un computador y de un punto de acceso.

Análisis e investigaciones sobre brecha digital, reflejan resultados que se da debido a múltiples determinantes, que hacen posible el uso de la tecnología, autores como (Taylor, Zhu, Dekkers, & Marshall, 2003), asocian a los factores demográficos y socio-económicos como principales determinantes del acceso a internet, dándonos una perspectiva de cómo y en que están empleando los usuarios la tecnología, estos factores a su vez se relacionan con variables tales como: los ingresos de la familia, la raza, niveles de educación, el género.

2.3. Factores que generan la brecha digital

En la actualidad hay varios estudios que se han realizado sobre las desigualdades digitales, estos señalan que la desigualdad se origina a partir de diferentes factores, que se han convertido en las principales determinantes de la brecha digital, los que mayor incidencia han tenido son los ingresos y la educación, relacionándose con otras variables como son el entorno familiar y social. (DiMaggio et al., 2004), así al hablar de la desigualdad digital nos estamos refiriendo a los diferentes factores que pueden ser requeridos para hacer uso de la tecnología. Por su lado,(DiMaggio & Hargittai, 2001) sugieren cinco dimensiones en las que la brecha digital puede existir:

- Medios técnicos: (software, hardware, la calidad de la conectividad).
- Autonomía de uso: (lugar de acceso, la libertad de uso sin restricciones económicas o sociopolíticas).
- Los patrones de uso: (los tipos de usos de Internet).

- Las redes de apoyo social: (disponibilidad de pares a quien se pueda recurrir para obtener ayuda con el uso, el tamaño de las redes para fomentar el uso).
- La habilidad: (la competencia para utilizar los medios digitales de manera eficaz).

Dentro de los factores que ocasionan la brecha digital diferentes autores hacen hincapié en los siguientes:

- **Ingresos económicos**

Según. (Jansen & Fellow, 2010), el principal factor que ocasiona la brecha digital en países desarrollados como EE UU, son las diferencias en los niveles de acceso determinadas por los ingresos económicos, en los hogares con mayores ingresos el 96% accede al internet, superando así a los hogares de menores ingresos donde solo el 63% hace uso de este recurso. Así las posibilidades de contar con un equipo y una conexión, ocasiona uno de los principales factores sobre la brecha digital incluso en países desarrollados (Taylor, Zhu, Dekkers, & Marshall, 2003) en su estudio Central Queensland, sobre comportamientos de consumo de la TIC, los sistemas informáticos de la comunidad, concluyen que los ingresos mayores y menores corresponden con el ingreso a internet.

Sin duda los ingresos económicos son un factor importante permitiéndoles a países desarrollados y sub desarrollados hacer uso de las TIC mediante la obtención de tecnología, donde este factor se puede generar en una variable dependiente así como independiente permitiendo el desarrollo de los mismos.

- **Educación**

La educación es un factor que determina el nivel y calidad de uso que hacen los usuarios de la tecnología, esto se refleja en las necesidades de información que varían según el perfil educativo de los usuarios. Según (Gutierrez & Gamboa, 2010), este factor tiene una importante relación con los ingresos esto se evidencia al ver que personas con mayores ingresos tienden a tener un nivel de educación mejor y por ende hacen mejor uso de la tecnología siendo esta una herramienta fundamental para su desarrollo. Así estudios relacionados demuestran estas diferencias reflejadas en el nivel de educación, los usuarios con niveles más altos de educación hacen mejor uso del tiempo de conexión, y de las diferentes herramientas tecnológicas, a diferencia de usuarios con niveles bajos de

educación, reportando mayores beneficios para los primeros, (Van Dijk, 2006). La brecha digital como la brecha cognitiva, observadas en los principales ámbitos constitutivos del conocimiento se refiere principalmente a la educación y otros factores. (UNESCO, 2005). Los reportes del proyecto Pew Research Center (2010) arrojan datos que demuestran una evidente desigualdad en el acceso dependiendo del nivel educativo. Los niveles de acceso llegan al 89% y al 93% para quienes tienen un título universitario de pregrado y posgrado respectivamente, con una evidente diferencia con el 40% de quienes no alcanzaron un título de bachillerato.

- **Género**

Las diferencias entre ser hombre o mujer es un hecho determinante en el acceso a la tecnología la misma que se está estudiando dentro de la brecha digital, en la actualidad es el factor con menor incidencia, estudios demuestran que aún se refleja diferencias de género en el grado de desigualdad en el uso del internet. Según el estudio de (Castaño C. , 2008), solo el 35% de las mujeres jóvenes comprendidas entre las edades de 16-24 años que son usuarias de internet tienen un nivel alto de habilidades en el uso, frente al 48% de los hombres, siendo muy baja su participación en profesiones que hacen uso de la tecnología con un 0,7% frente al 2,6% de varones. Estudios realizados por la, (UNESCO, 2005). Hacen referencia a la población existente en países en vías de desarrollo el mayor porcentaje de analfabetos son mujeres donde una de dos no sabe leer, a diferencia de los países industrializados donde un número importante de mujeres son usuarias de internet, haciendo reseña a las diferentes desventajas que les impide acceder a las nuevas tecnologías. En el proyecto SIGTIC realizado por (Castaño, Martín, & Martínez, 2011), determina la aparición de una segunda brecha digital de género que no está determinada por el nivel de acceso, sino por el nivel de inclusión, donde refleja una importante diferencia de desventaja entre hombres y mujeres donde el uso de las TIC está clara en este aspecto, la intensidad y la frecuencia de uso y la habilidad del usuario difieren del género (Hargittai & Shafer, 2006), que determina los sitios que visitan tanto mujeres como hombres, es el contar con una familia y con determinado número de hijos; el tiempo en línea de las mujeres con hijos es esto es ocasionado por el conjunto de tareas específicas dentro del hogar que muchas veces son asumidas por la mujer.

- **Edad**

La brecha digital por tema de edad, presenta algunas similitudes con la de género, vemos que las poblaciones más jóvenes tienen un nivel más elevado de los usos de la tecnología en la actualidad, pero son parte del público más vulnerable ante las dificultades económicas y sociales (Hargittai et al., 2006), a diferencia de las personas adultas que tienen dificultades en el aprendizaje, ante el acelerado crecimiento de la tecnología, tornándose un obstáculo insuperable para esta parte de la población.

2.4. Tipos de brecha digital

La brecha digital como se ha venido mencionando hace referencia a las distinciones de acceso, lo que origina los diferentes tipos de brecha digital según el usuario y sus condiciones para acceder al uso de la tecnología, siendo tema de investigación de gran importancia, en el siguiente apartado se detalla cada uno de los tipos más relevantes de lo que es hoy por hoy la brecha digital:

2.4.1. Brecha digital de conectividad

Brecha de conectividad se asocia a la infraestructura, la que hace posible la conexión física a internet de los diferentes usuarios, en los cuales se consideran dos grupos y dos tipos de acceso en los grupos de usuarios los que acceden y los que no acceden a Internet, y para los tipos de acceso, se presenta el acceso formal a Internet y acceso real a Internet (Castaño J. , 2011), señala que el acceso formal a internet, es la delimitante a la disponibilidad de las infraestructuras de conexión necesarias para el acceso a internet, como primera estructura necesaria, hace referencia a los dispositivos, equipos y aparatos electrónicos que permiten acceder a conexión de internet, segunda estructura necesaria los conductos, tales como cableados, fibras, enlaces, etc., donde la segunda estructura para hacer uso y tener conexión a internet se debe pagar. Lo que respecta al acceso real a internet, este tipo de acceso se aparta un poco de la tecnología en sí, acercándose más a los individuos como usuarios, quienes por el hecho de tener acceso y conexión a internet, no precisa de que estén haciendo uso del mismo, así lo determina el autor (Castaño J. ,

2011), en el mismo estudio, acceso real a internet, se ve condicionado a que la conexión a internet puede existir pero los usuarios pueden optar por no hacer uso, generando así el segundo aspecto de brecha digital.

2.4.2. Brecha digital de uso

La existencia de la brecha de uso, se fundamenta en la capacidad o la dificultad de desarrollar destrezas de una persona para hacer uso de la tecnología, basándose en las diferencias entre tener acceso y hacer buen uso de las TIC. Según, (DiMaggio et al., 2004), este tipo de brecha se enfoca en las distintas formas de utilizar la tecnología que hacen los usuarios la cual asume que los ingresos proveen un acceso privilegiado y la educación provee la perspicacia privilegiada. (Graham , 2008), en su investigación relaciona a la brecha digital de uso con el conocimiento, donde grupos sociales de clase alta que tienen mayores ingresos y mayores niveles de educación, permitiéndoles hacer un mejor uso del internet y con mayor frecuencia, obteniendo mejores beneficios que grupos sociales de clase media y la parte trabajadora. (Hargittai, 2010), descubrió que a pesar de contar con equipos y acceso, los estudiantes universitarios realizaban actividades diferenciadas en el internet, las cuales estaban determinadas con el nivel socio-económico de donde provenían. Para (Castaño Muñoz, 2010) ,la alfabetización digital juega un papel muy importante a la hora de adquirir habilidades en el uso de la tecnología. Las cuales están condicionadas por diferentes factores, como el ambiente donde se desarrollan las personas, el género, la clase de educación formal. Los estudiantes universitarios son un ejemplo, por el tipo de actividades de índole académico, que estos realizan están sujetos hacer uso del internet, estas son diferentes a las actividades que lleva un profesional desde su lugar de trabajo, y difiere también de los sitios web que visita una ama de casa. Biendo reflejado el tipo de uso y el nivel de uso que se le da a la tecnología basada en diferentes perfiles de usuarios, según sus habilidades, necesidades e intensidad de uso. La brecha digital de uso no es más que obtener, hacer y la intensidad del uso que podamos darle al internet para aprovechar mejor y obtener mejores beneficios de las diferentes herramientas que nos ofrece.

2.4.3. Brecha digital conceptual

La coexistencia de esta brecha hace referencia a usuarios de edades avanzadas que no conocen los beneficios de las TIC y que consideran que no necesitan hacer uso de ella, esto debido a que internet llegó a su apogeo mucho después, cuando estos usuarios ya tenían un estilo de vida establecido, padeciendo cierta fobia a la utilización de la misma basándose en predicciones sociales, culturales y psicológicas (Gutiérrez et al., 2010). Si hacemos referencia a este tipo de brecha en estudiantes universitarios no forman parte esta clasificación (Castaño et al., 2010), esto se argumenta a la naturaleza de estos usuarios que tiene la necesidad de hacer uso de las herramientas que ofrece el internet. Según (INEC, 2013), en el Ecuador los estudiantes universitarios no son parte de la brecha digital conceptual, estos cuentan con acceso a internet ya sea provisto por su institución educativa, desde el hogar, *telefonía móvil*, o desde un *cyber café*, el cual en 31,7% lo utiliza como medio de educación y aprendizaje, como parte de sus actividades académicas y la necesidad de contar con destrezas del manejo de la tecnología, lo que los convierte en usuarios frecuentes, haciendo uso de la tecnología en su rutina diaria.

2.5. Brecha digital en Latinoamérica

Al investigar el alcance de la brecha digital en América Latina se evidencia que aún queda un largo camino por recorrer en la erradicación, según el Informe de (SITEAL, 2012), Sistema de Información de Tendencias Educativas en América Latina, sobre brecha digital la población que cuenta con internet en el hogar, de 13 países analizados solo un 18% tiene acceso a internet desde su hogar, donde al menos 8 de 10 personas no cuenta con este recurso, observando que en América Latina, las oportunidades de acceso son muy desiguales entre países con porcentajes bastante elevados entre los mismos:

Tabla 1: Población de América Latina que cuenta con acceso a internet desde su hogar (en %) en el 2000

País	Personas que cuenta con internet (en %) en la década del 2000
Bolivia	3,6
Brasil	24,5
Chile	19,2
Colombia	12,0
Costa Rica	19,0
Ecuador	7,4
El Salvador	4,2
Guatemala	1,5
Honduras	2,4
Panamá	9,3
Paraguay	0,3
Perú	11,0
Uruguay	30,6

Fuente: Informe de SITEAL

Elaboración: Propia

A partir de la información presentada, se puede identificar que según el porcentaje de acceso a internet desde su hogar, se forman dos grupos de países, donde el primero oscila en porcentajes hasta el 5% demostrando que este recurso se extiende débilmente entre su población, y un segundo grupo con el 19 % y hasta el 30% determinado que aunque el acceso a internet de su población es elevado, aún queda mucho para que se llegue a hacer una distribución equitativa.

Por otro lado el foro (UIT, 2012), uno de los organismos más importantes, acerca del crecimiento de las TIC's en el mundo, en su edición del 2012, destacó los principales datos estadísticos de América Latina y el Caribe.

- En 2011, el promedio de penetración móvil-celular superó el umbral del 100% en América Latina y el Caribe.

- A finales de 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 ancho de banda internacional de Internet se multiplicó por doce en los últimos cinco años en América Latina y el Caribe, alcanzando los 4 700 Gbit/s a finales de 2011. Esto corresponde a más de diez veces del ancho de banda internacional de Internet en África.
- 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.

En América Latina dentro del ranking 2014 publicado por la UIT, de un total de 148 países a nivel mundial, podemos evidenciar las variantes que han tenido los países latino-americanos, donde el mejor posicionado es Panamá que ocupa el ranking 43, así como países que han sabido subir varias casillas entre ellos Perú que ha subido 13 casillas durante estos años, y otros en descenso en los cuales se encuentra México y Brasil quienes cayeron 16 y 9 casillas respectivamente.

2.6. Brecha digital en Ecuador

Informes de diferentes organismo internacionales que se encargan de medir el uso de las TIC's a nivel mundial y nacional, nos permiten tener una perspectiva de la conectividad de los consumidores, quienes tienen la necesidad de permanecer el mayor tiempo conectados a la red, desde diferentes puntos de conexión, dispositivos móviles, computadores personales, hogar, institución educativa, lugar de trabajo, sitios de entretenimiento, etc. Ecuador no es la excepción, observando que en los últimos años ha logrado significativos

adelantos en lo que respecta a tecnología, a niveles de acceso y conexión. Dentro del foro internacional publicado por la UIT podemos evidenciar que Ecuador ha ido en ascenso subiendo 9 casillas hasta el año 2014 dentro del ranking, demostrando con estos valores que como país está tratando de superar la brecha digital que lo separa de los países desarrollados. Por otra parte el informe (INEC, 2013), uno de los organismos dedicados a medir el uso de las TIC en el Ecuador hace referencia de la penetración en la utilización de internet, donde el 40% de la población ha utilizado internet (se ha recolectado los datos más notables del estudio):

- De las personas que usan Internet, el 45,1% lo hace en su hogar. En el área urbana el mayor porcentaje de la población utiliza Internet en el hogar con el 50,9%, mientras el mayor porcentaje de población del área rural lo usa en centros de acceso público con el 42,5%.
- En el 2013, el 32,0% de las personas usó Internet como fuente de información, mientras el 31,7% lo utilizó como medio de educación y aprendizaje.
- El 64,0% de las personas que usa Internet lo hacen por lo menos una vez al día, seguidos de los que por lo menos lo utilizan una vez a la semana con el 32,7%.
- El grupo etario con mayor uso de teléfono celular activado es la población que se encuentra entre 25 y 34 años con el 76,5%, seguido de los de 35 a 44 años con el 76,0%.
- La provincia con mayor porcentaje de personas que tienen celular activado es Pichincha con el 60,9%, seguida de Guayas con el 54,2%. Mientras que Chimborazo con el 37,4% es la de menor porcentaje.
- El 16,9% de las personas que posee un celular tiene un teléfono inteligente (*Smartphone*), frente al 8,4% del 2011, es decir 8,5 puntos más para el año 2013.
- En el 2013, el 20,0% de las personas en el Ecuador son analfabetas digitales, 9,2 puntos menos que en el 2010.

2.7. Redes sociales en la educación

Las redes sociales en internet se han ganado un lugar de una manera tan rápida convirtiéndose en prometedores negocios para empresas y sobretodo en lugares para la interacción entre humanos ya sean conocidos o desconocidos. Para comprender mejor de

que se trata este fenómeno que crece a pasos agigantado en internet, también se está haciendo uso de estas herramientas en la educación.

2.8. Internet y la educación superior

Con la incorporación de las nuevas tecnologías de la información y comunicación, y los innumerables recursos que nos ofrece para dar apoyo a la enseñanza aprendizaje, haciendo los espacios de educación más flexibles y adaptables tanto para educadores como para educandos de nivel superior. Donde el rendimiento académico hace referencia a la evaluación de los conocimientos adquiridos en el ámbito académico. Un estudiante con buen rendimiento académico es aquél que obtiene calificaciones positivas en los exámenes que debe rendir a lo largo de una carrera. Pero este rendimiento está relacionado a distintas determinantes. Según: (Garbanzo Vargas, 2007) Se agrupan en tres categorías: determinantes personales, determinantes sociales y determinantes institucionales que, a su vez, poseen sus propios indicadores.

Así podemos ver que la conceptualización del rendimiento académico ha variado a lo largo de la historia desde definiciones que se basan en un único criterio hasta las más actuales concepciones multidimensionales (Fullana, 1992). Teniendo que explicarse por la interacción de múltiples variables que forman una compleja red difícil de analizar, entre estas variables tenemos las personales, sociales y escolares. Según (Villa & Poblete, 2007), en estos últimos años la educación está basada en competencias las mismas que se ha visto conveniente crearlas para mejorar la calidad de la educación de los estudiantes estas competencias están comprendidas entre competencias genéricas instrumentales, genéricas interpersonales y competencias genéricas sistemáticas, las mismas que evalúan el perfil del estudiante para de esta forma poder inculcar al alumno que el éxito en este mundo globalizado y cambiante está en aprender a aprender.

2.9. Minería de datos

Hoy en día el bien más importante que tienen las empresas e instituciones de índole públicas, privadas o gubernamentales es la información, la cual esta almacenada en sistemas informáticos en grandes cantidades los cuales deben hacer proceso minuciosos

para obtener información que permita hacer una buena toma de decisiones. (Hernández, Ramírez, & Ferri, 2007), afirman que los datos pasaron de ser el resultado histórico de los sistemas de información denominado “producto” a ser una materia prima la cual hay que explotar para obtener lo esencial “producto elaborado” denominado conocimiento. Es así como (Pérez & Santín, 2007), puntualizan que la minería de datos es un proceso de descubrimiento de nuevas y significativas relaciones, patrones y tendencias para analizar grandes cantidades de datos.

Es así como la minería de datos no es más la encargada de realizar procesos mediante métodos que permitan procesar grandes cantidades de información, en niveles desconocidos y en formatos diferentes con la finalidad de extraer conocimiento útil y comprensible.

2.9.1. Tipos de datos

Dentro de la minería de datos los tipos son aplicables a cualquier dato, la variante se ve en las técnicas de minería, donde depende del tipo de dato para su aplicabilidad, en minería reconoce a los siguientes tipos de datos:

2.9.1.1. Bases de datos relacionales

Como su nombre los dice son relacionales y su principal característica es almacenar los datos de una forma ordenada y estructurada, obteniendo datos a través de lenguajes de consulta diseñados especialmente para esto.

(Hernández et al.,2007). Afirma que una base de datos relacional es una colección de relaciones (tablas), las cuales cada una consta de un conjunto de atributos (columnas o campos) y puede contener un gran número de tuplas (registros o filas). Cada tupla representa un objeto, el cual se describe a través de los valores de sus atributos y se caracteriza por tener una clave única y primaria que lo identifique.

2.9.1.2 Otros tipos de bases de datos

A pesar de que las bases relacionales como por ejemplo: My SQL Oracle, son las más utilizadas hoy en día existen aplicaciones que requieren de otro tipo de organización de la información, entre estas tenemos las de tipos de datos complejos como son:

- **Las bases de datos espaciales o geográficas.**- Contienen información relacionada con datos geográficos de gran volumen, como mapas, imágenes médicas, redes de transporte o tráfico, etc., siendo sus relaciones muy relevantes y de amplio espacio, de una área y materia en particular.
- **Las bases de datos temporales.**- La mayoría de su información está relacionada con atributos del tiempo, estos puede referirse a: distintos instantes o intervalos.
- **Las bases de datos documentales.**- Contienen descripciones para los objetos, que constan de documentos no estructurados como (biblioteca digital de novelas), semi-estructurados, donde se puede extraer información por partes (índices, resúmenes) o estructurados (como bases de datos de fichas bibliográficas).
- **Las bases de datos multimedia.**- Soporta almacenar objetos de gran tamaño como: imágenes, audio, videos que necesitan varios gigabytes de almacenamiento.

2.9.1.3 La World Wide Web

Está catalogada como el repositorio de información más grande y diversa de la actualidad, de la que podemos extraer conocimiento relevante y útil. Pero una gran desventaja es que la mayoría de información no está estructurada o semi-estructurada, dificultando así la minería web que es la que determina a qué páginas se debe acceder, para esto se organiza en tres categorías que son:

- Minería de contenido.- Para encontrar patrones de las páginas web.
- Minería de la estructura.- Entendiendo por estructura los hipervínculos y URLs.
- Minería de uso.- Que hace el usuario de las páginas web (navegación).

2.9.2. Tipos de modelos

La minería de datos ve de vital importancia hacer un análisis de los datos para extraer el conocimiento, que puede ser de forma de relaciones, patrones o reglas o un resumen a través de modelos, estos modelos son de dos tipos predictivos y descriptivos:

- Modelos predictivos.- Estiman valores futuros o desconocidos de variables, usando otras variables o campos de la base de datos como variables independientes o predictivas. Cuyos resultados son denominados variables objetivo o dependiente.
- Modelos descriptivo.- Sirven para explorar propiedades de los datos examinados a través de patrones que los explican y resumen.

2.9.3. La minería de datos y el proceso de descubrimiento de conocimientos en bases de datos

Según (Hernández et al., 2007), la minería de datos es la etapa de análisis de la información para la extracción o descubrimiento de conocimiento en bases de datos. Mientras que KDD, Descubrimiento de Conocimiento a partir de Bases de Datos por sus siglas en inglés (KDD, Knowledge Discovery from Databases), se define como el proceso no trivial de identificar patrones válidos, novedosos y potencialmente útiles y en última instancia, comprensible a partir de los datos. De estos términos parte la relación entre KDD y minería de datos: el KDD es el proceso completo de descubrir conocimientos útiles desde bases de datos, mientras que la minería de datos es una fase de KDD que hace referencia a la aplicación de los métodos de aprendizaje y estadísticos para la obtención de patrones y modelos.

No es un proceso automático, se caracteriza por ser iterativo e interactivo, iterativo con la finalidad de poder regresar de una fase a otra las veces que sea necesaria, para verificar resultados, con la finalidad de extraer información de alta calidad, interactivo ya que requiere de un especialista para hacer la validación y extracción del conocimiento. La siguiente figura ilustra las etapas del proceso KDD:

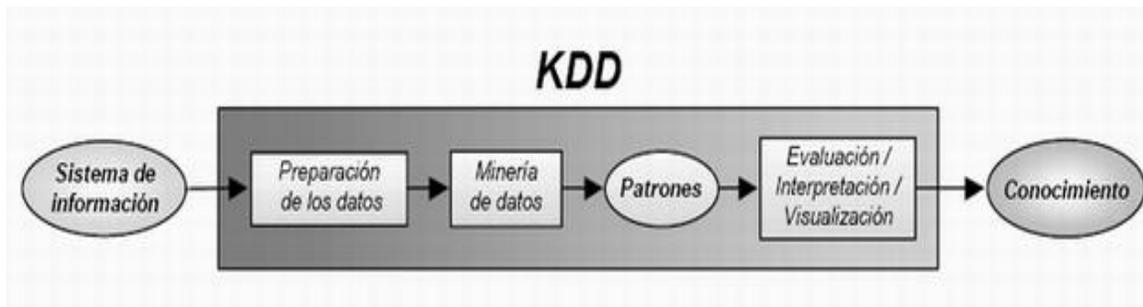


Figura 1: Proceso de KDD
Fuente: (Hernández et al., 2007)

2.9.3.1. Fase de integración y recopilación de datos

Esta fase es la encargada de recopilar la información necesaria de diferentes bases de datos, estas pueden ser de tipo públicas o privadas, respetando siempre las políticas de legibilidad para la extracción de información, para luego hacer una integración de estos datos formando un solo almacén de datos, con la finalidad de dejarlos accesibles para el análisis y la toma de decisiones, mediante una estructura multidimensional formada por atributos.

2.9.3.2. Fase de selección limpieza y transformación

El objetivo de esta fase escoge de forma coordinada los datos más relevantes, pues se deben eliminar el mayor número posible de datos erróneos o inconsistentes (limpieza) e irrelevantes (criba), minando los datos para mejorar su calidad, para obtener conocimientos relevantes al final del proceso.

2.9.3.3. Fase de minería de datos

Esta fase es la más precedente dentro del proceso KDD, que tiene como principal objetivo la transformación de la información recolectada en fases anteriores en nuevo conocimiento que sea accesible al usuario. La fase de minería de datos realiza este proceso mediante el empleo de algoritmos, métodos, técnicas, todo esto a través de tareas.

- **Tareas de minería de datos**

Afirma (Dialnet, 2010), Las tareas son actividades que dentro de la minería de datos son resueltas por algoritmos dependiendo del tipo de problema. Se diferencia dos grupos; en el primer se encuentran las tareas predictivas las que se encargan de predecir valores de interés a partir de otros valores que reflejen la generalidad de los datos, en este primer grupo de tareas predictivas encontramos la clasificación y regresión, en el segundo grupo se encuentran las descriptivas, que se encargan de identificar patrones en los datos para explicarlos o resumirlos, en este grupo podemos definir varias tareas descriptivas como: agrupamiento (*clustering*), reglas de asociación, reglas de asociación secuencial y correlaciones.

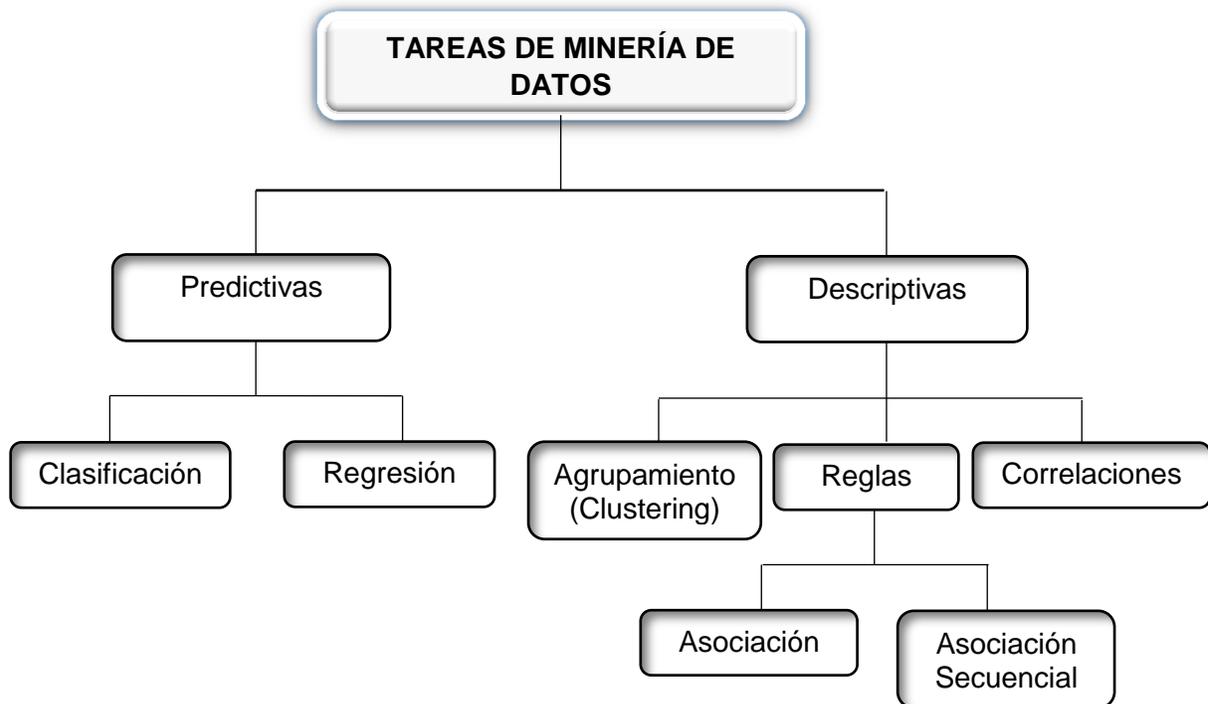


Figura 2: Representación de las Tareas de Minería de Datos
Fuente: Elaboración propia

- Tareas Predictivas (métodos supervisados)

La clasificación: consiste en encontrar tanto valores y variables significativas de las instancias de cada tipo de clase, para explorar las características de esta instancia y asignarle una clase predefinida.

Regresión: minimiza los errores entre el valor predicho y el valor real, adquiriendo conocimiento de funciones reales para asignarle un valor real que siempre será numérico.

- Tareas descriptivas (métodos no supervisadas)

Agrupamiento (*clustering*): tiene el propósito de formar grupos naturales homogéneos, desde grupos diversos, basándose en el grado de semejanzas entre ellos. Estos grupos son formados basándose en el principio de maximizar las similitudes de entre los elementos de un grupo y minimizar las similitudes entre los distintos grupos.

Correlaciones y factorización: determina el nivel de similitudes de los valores de dos variables numéricas. El objetivo de esta tarea es ver la relevancia de atributos, detectar atributos redundantes o dependencias entre atributos. Los estudios correlacionales y factoriales se centran exclusivamente en los atributos numéricos y permiten ver si dos variables están correlacionados linealmente o relacionados de algún otro modo (Hasperué, 2012)

Reglas de asociación: Esta tarea es muy similar a la tarea de correlación y factorización, su principal objetivo es identificar relaciones no explícitas entre atributos categóricos.

Reglas de asociación secuenciales: es utilizada para buscar patrones secuenciales dentro de los datos, basados en secuencias temporales y difieren de las reglas de asociación.

- **Técnicas de minería de datos**

Según (Moreno, 2007), para dar solución a problemas se hace uso de tareas como las descritas en la sección anterior, siendo requerido hacer uso de técnicas las cuales están basadas en métodos, estrategias o algoritmos que se aplican sobre conjunto de datos para obtener resultados.

Las técnicas más representativas son:

- Redes neuronales

En la actualidad es una de las técnicas más aplicadas para la detección de categorías comunes de entre los datos, debido a su capacidad de detectar y aprender complejos patrones, y características de los datos. Siendo característica de esta técnica de trabajar con

datos incompletos que a veces puede resultar en ventaja o inconveniente, es de dos tipos supervisado y no supervisado.

- Árboles de decisión

Es un modelos de predicción, como su nombre lo dice es representado en forma de árbol en donde cada nodo es una decisión, los cuales a su vez generan reglas para la clasificación de un conjunto de datos. Los árboles de decisión son fáciles de usar, admiten atributos discretos y continuos, tratan bien los atributos no significativos y los valores faltantes. Su principal ventaja es la facilidad de interpretación.

- Algoritmos genético

Los algoritmos genéticos imitan la evolución de las especies mediante la mutación, reproducción y selección, como también proporcionan programas y optimizaciones que pueden ser usadas en la construcción y entrenamiento de otras estructuras como es el caso de las redes neuronales. Además los algoritmos genéticos son inspirados en el principio de la supervivencia de los más aptos.

- Aprendizaje automático

Esta técnica de inteligencia artificial es utilizada para inferir conocimiento del resultado de la aplicación de alguna de las otras técnicas antes mencionadas.

2.9.4. Reconocimientos de patrones

Según, (Carrasco & Martínez, 2011). El reconocimiento de patronos es una interpretación del mundo, considerada también como la ciencia que se encarga de los procesos sobre: ingeniería, computación y matemáticas concernientes con cuerpos físicos o abstractos, con el propósito de extraer información que permita establecer propiedades de entre conjuntos de dichos objetos.

A continuación se ilustra el proceso de reconocimientos de patrones:

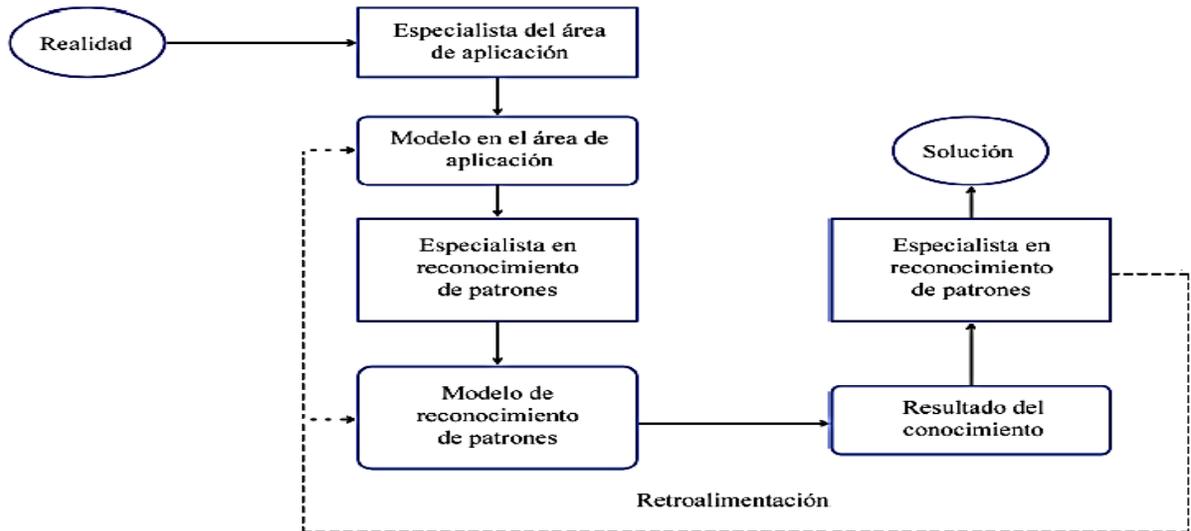


Figura 3: Proceso del Reconocimiento de Patrones
Fuente: (Carrasco & Martínez, 2011)

2.9.5. Clustering

Como se mencionó anteriormente, *Clustering* análisis es el estudio formal de métodos y algoritmos para la agrupación de los objetos de acuerdo con la medida o percibidas características o similitud intrínsecas, (Jain, 2009).

(Cardenas, 2013) Afirma que *Clustering* es el proceso de agrupar los datos en clases o en clústeres, de tal forma que, los datos de un mismo clúster tienen una alta similitud y a su vez, son muy diferentes de los de otro clúster. Un clúster de objetos puede ser tratado colectivamente como un grupo o ser considerado como una forma de compresión de datos. El proceso de agrupamiento de *clustering* es desarrollado por algoritmos el más popular y simple es:

K-means: dentro de del análisis clúster o agrupamiento es considerado uno de los algoritmos más populares, por ser iterativo eficaz y fácil de utilizar para hacer la división de una base de datos en k-grupos, uniendo los centroides y luego tomar cada punto de la base de datos y situarlo en la clase de su centroide más cercano, este proceso se repite hasta que ya no existan cambios entre los grupos de un paso al siguiente.

CAPITULO III: METODOLOGÍA

Este capítulo se hace referencia del proceso metodológico que se usó para la alcanzar los objetivos planteados en el presente proyecto investigativo, los usos que le están dando a la tecnología los estudiantes de las universidades ecuatorianas. De igual forma se detallan los instrumentos de levantamiento de información y las técnicas estadísticas que se emplean en el desarrollo del presente proyecto.

3.1. Descripción de la población y tamaño de la muestra

El levantamiento de la información se la realizo en la Universidad Técnica del Norte, de la modalidad de estudios presencial, conformada por 5 facultades que ofertan 33 carreras, con un estimado de 7.500 estudiantes, universidad que es parte del proyecto de investigación “La educación virtual en el Ecuador”, conformado por 35 universidades a nivel de todo el Ecuador de las categorías A, B y C.

Para determinar el tamaño de la muestra final se aplicó la formula general para poblaciones finitas, citada por (Castellanos, 2009), la población universitaria estimada 7.500 estudiantes, valor para determinar el tamaño de la muestra, se ha utilizado la fórmula de cálculo $p = q = 50$, que permite maximizar el tamaño. El nivel de confianza es del 95% y se acepta un margen de error del 5%. El resultado de aplicar la formula nos arroja un valor de 400 estudiantes que deben encuestarse, sin embargo en la presente investigación se han encuestado a 500 estudiantes.

La fórmula aplicada se detalla a continuación:

$$n = \frac{N * Z_{\alpha}^2 * p * q}{d^2 * (N - 1) + Z_{\alpha}^2 * p * q}$$

- n = tamaño de la muestra
- N = Total de la población
- $Z^2 = 1.96$ al cuadrado (la seguridad es del 95%)
- p = proporción esperada (en este caso $5\% = 0.05$)
- $q = 1 - p$ (en este caso $1 - 0.05 = 0.95$)
- d = precisión (en la investigación se usó un 5%).

Despeje de la fórmula:

$$n = \frac{N * Z_{\alpha}^2 * p * q}{d^2 * (N - 1) + Z_{\alpha}^2 * p * q}$$

$$n = \frac{7.500 * 1.96^2 * 0.05 * 0.95}{0.05^2 * (7.500 - 1) + 1.96^2 * 0.05 * 0.95}$$

$$n = \frac{7500 * 3.8416 * 0.05 * 0.95}{0.0025 * (7499) + 3.8416 * 0.05 * 0.95}$$

$$n = \frac{7.500 * 0.182476}{18.7475 + 0.182476}$$

$$n = \frac{1368.57}{3.42097}$$

$$n = 400.0532$$

$$n = 400$$

3.2. Instrumentos de recolección de información

El levantamiento de la información a los estudiantes de la Universidad Técnica del Norte, se la realizó aplicando métodos de investigación tales como: encuestas y entrevistas, métodos que son más conocidos, de fácil aplicación y la obtención de la información es lo más precisa y veras:

- Encuesta: Las encuestas fueron aplicadas a 500 estudiantes desde segundo ciclo en delante de diferentes carreras, y luego de una limpieza de coherencia en los datos se descartó algunas de las mismas quedando una muestra de 452, la cual tiene como objetivo dar solución a las hipótesis de este trabajo investigativo, donde recoge información sociodemográfica y socioeconómica de los estudiantes, encuesta que fue

realizada por el director del proyecto “La educación virtual en el Ecuador” (Torres J. C., 2012), ver (Anexo 1).

- Entrevista: La entrevista se la realizo al Ing. Juan Carlos García, jefe de proyectos informáticos de la Universidad Técnica del Norte, quien tuvo la amabilidad de colaborarnos, entrevista que abarca temas como: instalaciones físicas y políticas de uso de la tecnología, dentro del campus universitario, ver (Anexo 2).

3.3. Métodos aplicados para el análisis e interpretación de datos

Para el proceso de análisis e interpretación de los datos se realizó mediante la metodología de KDD (Descubrimiento de Conocimiento en Bases de Datos), el cual nos permite hacer un análisis inteligente de los datos, haciendo hincapié en las técnicas de análisis estadísticas, el mismo que está organizado en torno a 5 fases como son:

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

Fases que trabajan para encontrar patrones que permitan describir la información que sea de fácil interpretación para los humanos, dadas las características del proyecto investigativo y el previo análisis de que realiza cada una de las fases, se aplicó a este proyecto hasta la fase 4, las cuales nos permitieron dar solución alcanzando los objetivos planteados.

A continuación se describen las fases de minería de datos aplicadas en el proyecto investigativo:

3.3.1. Fase de integración y recopilación de datos

La fase de integración y recopilación de los datos es la primera fase dentro del proceso de extracción del conocimiento, la que determina las fuentes de información y donde conseguirlas, la cual se aplicó para obtener una base de datos que nos permita seguir avanzando con el proyecto y la ejecución de las demás fases.

El levantamiento de información se la realizo, siguiendo los diferentes procesos, dirigiéndose al Rector de la Universidad Técnica del Norte mediante una carta (Anexo 3), para tener apertura e ingresar a las instalaciones de la universidad y aplicar las encuestas a los estudiantes (<https://www.youtube.com/watch?v=3loQGom0mHc>), de cada una de las facultades de diferentes carreras, y una entrevista a un directivo (<https://www.youtube.com/watch?v=qVG0IF6BckY>). Obteniendo como resultado una respuesta positiva y toda la colaboración por parte del Rector, sus Catedráticos y el Alumnado (Anexo 4).

La encuesta aplicada, está formulada por 17 preguntas las que nos permiten indagar información de los estudiantes, información requerida para el desarrollo del proyecto de investigación:

- Información académica del estudiante, que pone a consideración las preguntas 1 y 2.
 - En qué universidad estudia.
 - Qué carrera estudia.
- Información sociodemográfica, información concentrada, aplicando las siguientes variables:
 - Edad.
 - Género.
 - Niveles de ingresos familiares, medidos en una escala de 5 niveles.
- Preguntas que determinan el perfil de conexión a internet de los estudiantes (internauta). Recolectados en las preguntas de la 6 a la 9, con variables denominadas categóricas:
 - Lugar de conexión a internet, en un intervalo de 5 niveles (1: desde la casa, 2: desde un *cyber café*, 3: desde el trabajo, 4: desde la universidad, 5: desde una red móvil (*movistar, claro, cnt*)).
 - Niveles de conexión diaria a internet, en un intervalo de 7 niveles.
 - Grado de conocimiento en el manejo de internet, en un intervalo de 1 al 10.
 - Horas de conexión al día.
 - Años de conexión a internet.
- Actividad académica del estudiante medida semanal y mensualmente en la pregunta 10, conformada por 10 ítems.
- Uso del Internet para entretenimiento y diversión: agrupada en la pregunta 11, formada por 5 ítems, variables de tipo ordinal (Anexo 1).

- Perfil de uso de las redes sociales, concentrado en la pregunta 12, basada en el número de seguidores en las plataformas más comunes de redes sociales, como: (*Twitter, Facebook, LinkedIn*)
- Uso de la Web, la misma que considera tres variables con respuestas de tipo categórica. Variables agrupadas en la pregunta 13.
- Niveles de uso de dispositivos tecnológicos: donde se pone a consideración dispositivos móviles de uso personal, medidos en intervalos de 1 al 10, puestos a consideración en la pregunta 14 (7 ítems) (Anexo 1).
- Apreciación de los estudiantes del uso de tecnología para desarrollo de actividades académicas, medida en la pregunta 15, conformada por 6 ítems, en 10 niveles.
- Apreciaciones de los estudiantes sobre el uso de tecnología por parte del docente para impartir actividades de enseñanza-aprendizaje, esta información se la obtiene en 11 variables agrupadas en la pregunta 16 del (Anexo 1).
- Determinación del rendimiento académico: información recolectada en la pregunta 17, con dos variables de tipo numéricas.
 - En el semestre anterior en cuantas asignaturas se matriculo
 - En el semestre anterior cuantas asignaturas aprobó

Finalizada la aplicación de las encuestas se las enumero para proceder a subir manualmente en la herramienta *SurveyMonkey*, herramienta que permite crear y publicar encuestas online en minutos y ve los resultados gráficamente y en tiempo real, fueron subidas bajo la denominación del link: <https://es.surveymonkey.com/s/encuestaUTNorte>, el mismo que fue proporcionado por el director del proyecto (Torres J. C., 2012). Obteniendo de este proceso una base de datos para continuar con las siguientes fases:

3.3.2. Fase de selección, limpieza y transformación

Esta fase se encarga del control de la calidad de la información, después de hacer un análisis en la base de datos, eliminando o corrigiendo datos incorrectos existentes, para determinar qué estrategia seguir con los datos incompletos después de haber sido minados.

Proceso de limpieza de datos: en la fase anterior obtuvimos una base de datos con 500 registros, a la cual le aplicamos el proceso de limpieza de minería de datos, fase facultada para eliminar datos anómalos o valores atípicos, registros con datos faltantes o irrelevantes, registros con respuestas en blanco, registros con datos incoherentes, principales causantes de inconsistencia, análisis inexactos o incorrectos dentro de la base de datos, por ejemplo: estudiantes que no se matricularon en el semestre anterior pero que registraban una respuesta en las asignaturas aprobadas el semestre anterior y viceversa. Obteniendo al final una estructura de datos adecuada para su posterior transformación.

Proceso de transformación de datos: la transformación de los datos consistió en asignar a las variables existentes, etiquetas y tipos de datos, para generar nuevas variables, que contengan una estructura adecuada para una mejor identificación.

En la siguiente tabla se ilustra el proceso de transformación de variables:

Tabla 2: Variables determinadas transformación de los datos

Variable original	Descripción	Etiqueta final de la variable	Valor de la variable transformada en niveles	Asignación del tipo de dato	Medida determinada
Hombre Mujer	Genero del estudiante	Sex	1: Hombre 2: Mujer	Numérico	Dicotómica
Hasta 350 Hasta 600 Hasta 1000 Hasta 1500 Más de 1500 dólares	Ingresos familiares	ing	1 :Hasta 350 2: Hasta 600 3: Hasta 1000 4: Hasta 1500 5: Más de 1500 dólares	Numérico	Nominal
Desde la casa; Desde un cyber café Desde el trabajo Desde la universidad Desde una red móvil	Lugar habitual de conexión a internet	lu_con	1: Desde la casa; 2: Desde un cyber café 3: Desde el trabajo 4: Desde la universidad 5: Desde una red móvil	Numérico	Nominal
Aprobó No aprobó	Patrón aprobado o no aprobado por el estudiante		1: Aprobó 2: No aprobó	Numérico	Dicotómico

Fuente: Datos de la investigación

Elaboración: Propia

Una vez terminada la limpieza y transformación, mediante operaciones de discretización y numerización, con la finalidad de cambiar el tipo de datos de numérico a nominal o viceversa, se realizó un análisis para recalcar la concentración de las variables más relevantes para obtener información de la investigación, variables que hacen posible la comprobación de las hipótesis planteadas al iniciar este proyecto.

En la siguiente tabla se ilustra el proceso de selección de variables:

Tabla 3: Variables que intervienen en la comprobación de hipótesis

Grupo de origen	Variable para comprobar hipótesis	Variabes que intervienen en la clusterización	Variable que determinan el perfil del estudiantes	Hipótesis que comprobará	Tipo
Académicas	a_ing_pla a_con_pro a_con_est a_rec_edu a_vid_aca a_for_vir a_pos_aca a_hor_cha a_bus_inf a_bib_vir	a_con_pro a_con_est a_vid_aca a_hor_cha a_bus_inf	Perfil académico	H ₁	Dependiente
				H ₃	Independiente
Entretenimiento	e_hor_cha e_hor_red e_jue_lin e_des_con e_vid_sem	e_hor_cha e_hor_red e_jue_lin e_des_con e_vid_sem	Perfil de entretenimiento	H ₂	Dependiente
				H ₄	Independiente
Nivel de ingreso familiar	1:Hasta 350 2:Hasta 600 3:Hasta 1000 4:Hasta 1500 5:Más de 1500 dólares	N/A	Nivel de ingreso	H ₁ y H ₂	Independiente
Rendimiento académico	1: Aprobó 2: No aprobó	N/A	Rendimiento	H ₃ y H ₄	dependiente

Fuente: Datos de la investigación

Elaboración: Propia

3.3.3. Fase de Minería de datos

Esta fase tiene como objetivo analizar los datos para generar conocimiento que pueda ser utilizado por el usuario, partiendo de los datos recopilados. La aplicación de la fase de la minería de datos en este proyecto se la realizó centrándose en el objetivo principal que engloba esta investigación, establecer grupos o perfiles de estudiantes con características

común de uso de Internet; para proceder a la comprobación de las hipótesis planteadas al iniciar esta investigación.

Los pasos que se vieron apropiados seguir dentro de la fase minería de datos para el análisis y conformación de los perfiles de estudiantes fueron:

3.3.3.1. Análisis de centralización de los datos

- **Periodicidad de datos**

Dentro del proceso de periodicidad de los datos, se identificó la concentración de información en cada una de las preguntas y sus variables, para la obtención de datos estadísticos que se utilizaran en análisis posteriores.

- **Media:** suma de todos los valores en la población dividida entre el número de valores de la población.

$$\mu = \frac{\sum X}{N}$$

En la cual:

μ : representa la media poblacional; se trata de la letra minúscula griega mu;

N: es el número de valores en la población;

X: representa cualquier valor particular;

Σ : es la letra mayúscula griega sigma e indica la operación de suma;

ΣX : es la suma de X valores en la población.

- **Mediana:** Punto medio de los valores una vez que se han ordenado de menor a mayor o de mayor a menor.

$$Me = L_i + a \left(\frac{\frac{N}{2} - f_{i-1}}{f_i} \right)$$

En la cual:

L_i : Es el límite inferior de la clase mediana

$N/2$: Es la semisuma de las frecuencias absolutas

f_{i-1} : Es la frecuencia absoluta acumulada de la clase anterior a la mediana.

a_i : Es la amplitud del intervalo

Me: Mediana

- **Moda:** valor de la observación que aparece con mayor frecuencia. La moda en ocasiones puede no existir.

$$Mo = E_i \frac{d_i}{d_i + d_{i+1}} a_i$$

En la cual:

Mo: Representa la Moda

E_i : Es el límite inferior real de la clase modal

d_i : Representa la diferencia entre la frecuencia absoluta de la clase modal y la clase anterior.

d_{i+1} : Representa la diferencia entre la frecuencia absoluta de la clase modal y la siguiente.

a_i : Representa la amplitud del intervalo de la clase modal.

3.3.3.2. Análisis de correspondencia

Es una técnica descriptiva o explicativa que se aplica con la finalidad de resumir el volumen de datos en un número de dimensiones, sin perder mucha información, aplicada a variables categóricas u ordinales. En el análisis de correspondencia existen dos grupos; correspondencia simple y múltiples, (De la Fuente, 2011).

Una tabla de contingencia está condicionada por los siguientes datos para que pueda existir una correlación significativa:

- El nivel de significancia indica si existe o no relación entre las dos variables. Determinado por el valor menor de 0.05, para que exista correlación significativa.
- Chi-cuadrado; Medida que compara los valores (n_{ij}) observados en la tabla con los que teóricamente se obtendrían (t_{ij}) bajo la hipótesis nula, este estadístico se utiliza para identificar relaciones de dependencia.
- En escala nominal podemos considerar las siguientes medidas de asociación: Coeficiente ϕ , Coeficiente de contingencia o C de Pearson (C), Coeficiente, Coeficiente V de Cramer (V), Coeficiente Lambda (λ)

Dadas las características de la presente investigación, se realizaron tablas de contingencia entre las variables determinadas por su concentración de información para la comprobación

de las hipótesis. Se aplicó tablas de contingencia de análisis simple donde se obtuvieron los valores estadísticos de chi-cuadrado, R de Pearson y Tau-c, de acuerdo al tipo de variables y valores, ver (Anexo 5).

3.3.3.4. Reducción de variables

Dentro de la minería de datos tiene diferentes tareas para realizar la reducción de variables estas se aplican dependiendo de las características del proyecto.

A continuación se describen algunas de las tareas que fueron aplicadas:

- **Reducción de variables factorización**

Factorización tiene como objetivo reducir la dimensión de las variables, simplificándolas con la finalidad de hacerlas más fácil de interpretar. En presente proyecto de investigación tiene un amplio número de variables que están determinadas en medir el uso que le están dando los estudiantes al internet, en lo que respecta a lo académico y entretenimiento, para comprobar esto se ve necesario formar grupo homogéneos.

Para esto se realizaron varios proceso de factorización con distintas variables, pudiendo determinar que no existe la posibilidad de aplicar el proceso de esta tarea, puesto que el valor del determinante de la matriz de correlaciones es muy lejano a cero, lo que demuestra que no existe grado de certeza o validez.

Las condiciones que debe cumplir el análisis factorial para ser válido según (De la Fuente, 2011) son:

1. “Que el valor correspondiente a la determinante sea menor a cero ($d < 0$).
2. Que el valor de la prueba KMO sea cercano a 1.
3. Y el valor de la varianza total alcance un valor superior al 70% de exactitud”.

Los resultados obtenidos de la aplicación de la tarea de factorización se la puede, ver a detalle en los (Anexos 5 y 6)

- **Agrupación (*Clusterización*)**

El análisis clúster o conglomerados nos permite realizar una clasificación o agrupamiento conformada por n individuos en g grupos, determinado por el comportamiento de estos en un conjunto de variables. *Clúster* es una técnica multivariante que considera todas las variables como intervinientes concediéndoles a todas las mismas funciones en la configuración de los grupos. (Díaz de Rada, 2002). Para realizar un análisis clúster se establecen 4 etapas que son:

- Estudio pormenorizado de las variables que van a ser incluidas en el análisis.
- Uso de un método de clasificación de clúster con objeto de crear grupos similares.
- Cálculo de las distancias o similitudes entre los casos.
- Validación de los resultados del análisis.

En la investigación se aplicó el análisis clúster de clasificaciones no jerárquicas, que nos permite trabajar con muestras altas mediante la aplicación del algoritmo K-means.

- **Algoritmo K-means**

Es uno de los más simples y conocidos algoritmos de agrupamiento, sigue una forma fácil y simple para dividir una base de datos dada en k grupos, este algoritmo para realizar los grupos o conglomerados lo realiza mediante casos basándose en las distancias existentes entre los casos, (Díaz de Rada, 2002).

- Sitúa K puntos en el espacio en el que "viven" los objetos que se quieren clasificar. Estos puntos representan los centroides iniciales de los grupos.
- Asigna cada objeto al grupo que tiene el centroide más cercano.
- Tras haber asignado todos los objetos, recalcula las posiciones de los K centroides.
- Repite los pasos 2 y 3 hasta que los centroides se mantengan estables. Esto produce una clasificación de los objetos en grupos que permite dar una métrica entre ellos.

Mediante la aplicación del algoritmo k-medias, se pudo obtener la concentración de información en las variables de las preguntas 10 y 11 las cuales determinan los perfiles del uso del internet tanto en lo académico como para entretenimiento respectivamente.

- Perfiles de uso de internet en lo académico

En este grupo intervinieron las variables más representativas como: consultas a profesores, consultas a compañeros, videos académicos que mira en YouTube, chat sobre temas académicos, información académica que busca en internet, de las cuales se conformaron grupos de 2, 3, 4 y 5 seleccionando el grupo de 2 por su fácil representación; donde se los denomino regulares y dedicados. (Análisis de Resultados).

- Perfiles de uso de internet en entretenimiento

Para este perfil las variables más representativas son: chat por diversión, redes sociales, juegos en línea, descargas de música videos y programas, videos de entretenimiento que mira en YouTube. Así mismo se realizó la clasificación de 2, 3, 4 y 5 grupos, seleccionando el grupo 3, denominados intensivo, activo, pasivo. (Análisis de Resultados).

3.3.3.4. Discriminación (discriminante)

El Análisis Discriminante es una técnica estadística multivariante cuya finalidad es analizar si existen diferencias significativas entre grupos respecto a un conjunto de variables medidas sobre los mismos para que en caso de que existan, explicar en qué sentido se dan y facilitar procedimientos de clasificación sistemática de nuevas observaciones de origen desconocido en uno de los grupos analizados, (De la Fuente, 2011).

Conformados los grupos en la fase anterior con el algoritmo k-medias se deben comprobar, cuál de estos grupos tiene valores más representativos, mediante el análisis discriminante, podemos escoger la mejor clasificación.

Aplicado el análisis discriminante para los grupos 2, 3, 4 y 5, se escogió el clúster 2 para (perfiles de actividades académicas) tiene un porcentaje de exactitud de 99,30% como lo indica la tabla siguiente, para los proceso de los grupos 3, 4 y 5 podemos verlos en el Anexo 16.

Tabla 4: Discriminante de 2 grupos para actividades académicas

Resultados de clasificación^a					
		Número de caso de clúster	Pertenenencia a grupos pronosticada		Total
			1	2	
Original	Recuento	1	322	1	323
		2	2	127	129
	%	1	99.7	.3	100.0
		2	1.6	98.4	100.0
a. 99,30% de casos agrupados originales clasificados correctamente.					

Fuente: Datos de la investigación**Elaboración:** Propia

En lo que respecta al clúster de perfil para actividades de entretenimiento, aplicado el análisis discriminante para los grupos 2, 4 y 5, se obtuvo el valor más significativo en el grupo del clúster 3 con un porcentaje de exactitud del 97,60%, como podemos comprobar en la tabla (Tabla 5), los procesos del análisis discriminante para los otros grupos podemos ver en el Anexo 17.

Tabla 5: Discriminante de 3 grupos para actividades de entretenimiento

Resultados de clasificación^a						
		Número de caso	Pertenenencia a grupos pronosticada			Total
			1	2	3	
Original	Recuento	1	80	3	0	83
		2	1	93	3	97
		3	3	1	268	272
	%	1	96.4	3.6	.0	100.0
		2	1.0	95.9	3.1	100.0
		3	1.1	.4	98.5	100.0
a. 97,60% de casos agrupados originales clasificados correctamente.						

Fuente: Datos de la investigación**Elaboración:** Propia

3.3.4. Fase de evaluación e interpretación

En esta fase se hizo uso de los diferentes patrones para realizar la evaluación de los resultados obtenidos en la fase anterior mediante la comprobación de las hipótesis planteadas en esta investigación. Técnicas que fueron aplicadas para aceptar o rechazar las hipótesis, tales como: Regresión Logística (Pérez & Santín, 2007), es un modelo que permite verificar la probabilidad de ocurrencia de un evento, mediante la variable dependiente, es una de las técnicas más adecuada cuando se hace clasificaciones basadas en las características de los datos. (Porcel, Dopaso, & López, 2008).

Dentro de regresión logística existen dos grupos binomial y multinomial, las mismas que se diferencian en el número de categorías que obtenga la variable dependiente; es decir, será considerada 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 & San Martín Fernández, 2004).

3.3.4.1. Regresión logística multinomial

En la presente investigación se aplicó la regresión logística multinomial para dar contestación a la Hipótesis 2, debido a las características de las variables y funcionalidades de esta técnica que solo se aplica a variables dependientes con más de dos categorías; resultados que fueron tomados de la clusterización del perfil de entretenimiento que consta de 3 niveles y la variable independiente ingresos que tiene 5 niveles denominada como variable categórica nominal.

La técnica de regresión logística multinomial muestra las comprobaciones y ajustes de modelos mediante varios estadísticos. En la Tabla 6, se muestra los resultados de la aplicación de la técnica de Chi Cuadrado que se realizó para la comprobación de las relaciones significativas entre las dos variables, para verificar y dar contestación si se cumple o no la hipótesis planteada, verificando la relación significativa entre la variable independiente y la variable dependiente.

Tabla 6: Pruebas de Chi Cuadrado para la Hipótesis 2

Hipótesis	Nombre de la prueba	Valor	gl	Sig. asintótica (2 caras)
Hipótesis 2	Chi-cuadrado	21,059 ^a	8	0,007

Fuente: Datos de la investigación

Elaboración: Propia

3.3.4.1. Regresión logística binomial

Para continuar con el proceso de evaluación e interpretación también se hizo uso de la técnica de regresión logística binomial para verificar la comprobación de las Hipótesis 1, Hipótesis 3 e Hipótesis 4, estas hipótesis tienen la particularidad que la variable dependiente es una variable dicotómica, es decir que toman valores de cero "0" y uno "1". La Tabla 7, muestra los resultados de la prueba del Chi-cuadrado que fueron sometidas las variables de las Hipótesis 1, Hipótesis 3 e Hipótesis 4 para observar si existencia una relación entre las variables independientes con las variable dependientes.

Tabla 7: Pruebas de Chi Cuadrado para la Hipótesis 1, Hipótesis 3 e Hipótesis 4

Hipótesis	Nombre de la prueba	Valor	gl	Sig. asintótica (2 caras)
Hipótesis 1	Chi-cuadrado	1,288 ^a	4	0,863
Hipótesis 3	Chi-cuadrado	1,484 ^a	1	0,223
Hipótesis 4	Chi-cuadrado	1,233 ^a	2	0,540

Fuente: Datos de la investigación

Elaboración: Propia

CAPITULO IV: ANÁLISIS DE RESULTADOS

Dentro de este capítulo se recogen los resultados obtenidos mediante la aplicación de diferentes fases de minería de datos especialmente del proceso de KDD (Descubrimiento de Conocimiento a partir de Bases de Datos) de acuerdo con los objetivos, preguntas e hipótesis de investigación detallados en el capítulo I. Análisis que están descritos en el capítulo III que hace referencia a la metodología de investigación.

4.1. Particularidades de la población universitaria

4.1.1. Descripción de la muestra

En la Universidad Técnica del Norte la población estudiantil se estima en 7.500 estudiantes, valor al que se le aplico la fórmula de cálculo para poblaciones finitas, el resultado de aplicar dicha fórmula nos arroja, un valor de 400 encuestas que deben ser aplicadas, sin embargo en el presente proyecto investigativo se aplicaron 500 encuestas, las cuales de pasar por procedimiento de validación y concordancia de datos nos quedaron un valor de 452. Así podemos observar que de 452 participantes con 244 (54,0%) pertenecen al género masculino y 208 (46,0%) femenino con un porcentaje de diferencia 8% de esta forma podemos observar que la proporción de participantes masculinos es con muy poco mayor que la proporción de participantes del género femenino.

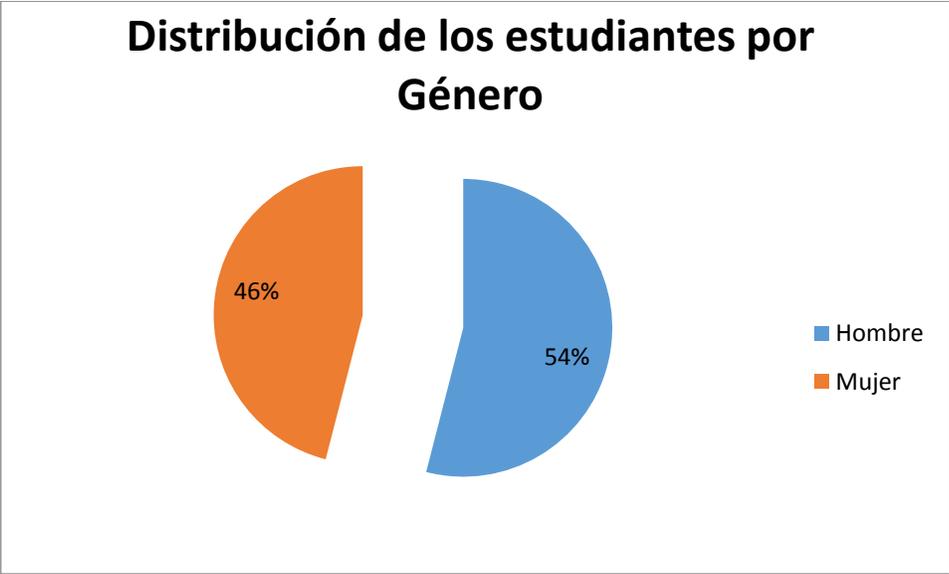


Figura 4: Distribución de los estudiantes por género
Fuente: Elaboración propia

En la Universidad Técnica del Norte se ofertan 33 carreras, pero el levantamiento de la información se lo realizó de 15 carreras, de las cuales se recolectó información para esta investigación por medio de sus estudiantes, podemos observar que en las carreras de Ingeniería en Turismo con un número de 69 (15,3%) y Licenciatura en Diseño Gráfico con 68 (15,0%) son las carreras con más número de alumnos.



Figura 5: Distribución de los estudiantes por titulación
Fuente: Elaboración propia

Las edades promedio de los estudiantes que nos colaboraron con la recolección de información oscilan de entre 18 a 36 años, como podemos constatar los que tiene una edad de 18 a 20 años tienen un porcentaje del 50,60% de un total del 100% con una desviación

estándar de 2,33% y una moda de 20. Así podemos constatar que la población estudiantil UTN, es bastante joven.

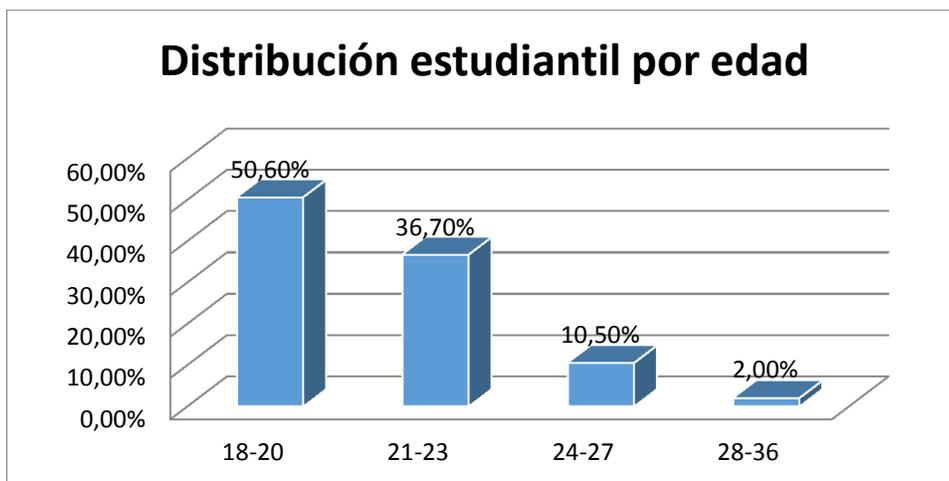


Figura 6: Distribución estudiantil por edad
Fuente: Elaboración propia

Los porcentajes de ingresos mensuales de las familias de los estudiantes encuestados se presentan en la figura 7, pudiendo constatar que los de mayor porcentaje son los sueldos de \$350,00 y \$600,00 dólares con el 34,7% y 33,2% respectivamente, de estos datos podemos determinar que aunque los ingresos familiares no son altos, no es un impedimento para optar por una formación de nivel superior.

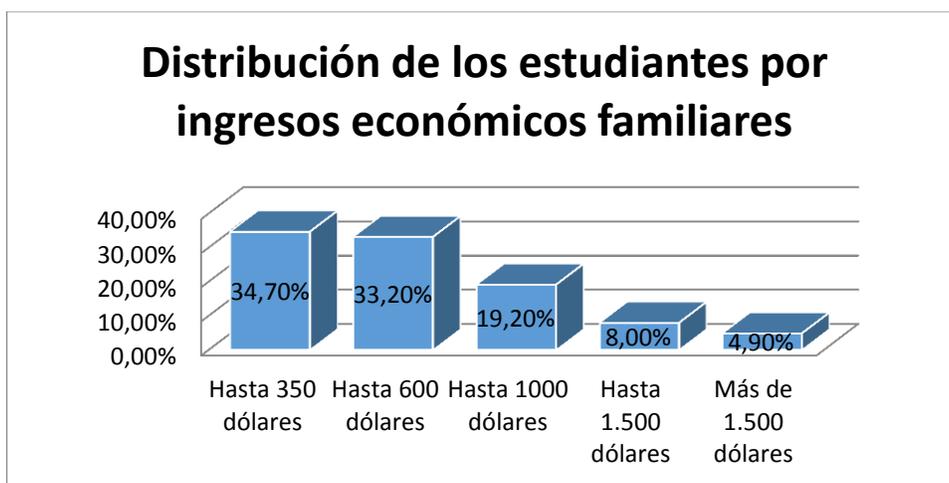


Figura 7: Distribución de los estudiantes por ingresos económicos familiares
Fuente: Elaboración propia

4.1.2. Preferencias de los estudiantes sobre uso del internet

En este apartado se presentan algunas de las distribuciones que permiten conocer las preferencias que tienen los estudiantes en la utilización del internet, delineando un perfil en función de sus posibilidades socio-económicas.

Se puede observar en la Figura 8, que el lugar habitual de conexión es desde la casa 73,50%, pero no determina 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 (0,40%). Se ha encontrado asociación significativa entre los ingresos y el lugar de conexión a internet que utilizan los estudiantes ($X_2 = 49,185$; $P=0,000$) el nivel de correlación señalado por V de Cramer es significativo y bajo ($V=0,165$; $P=0,000$).

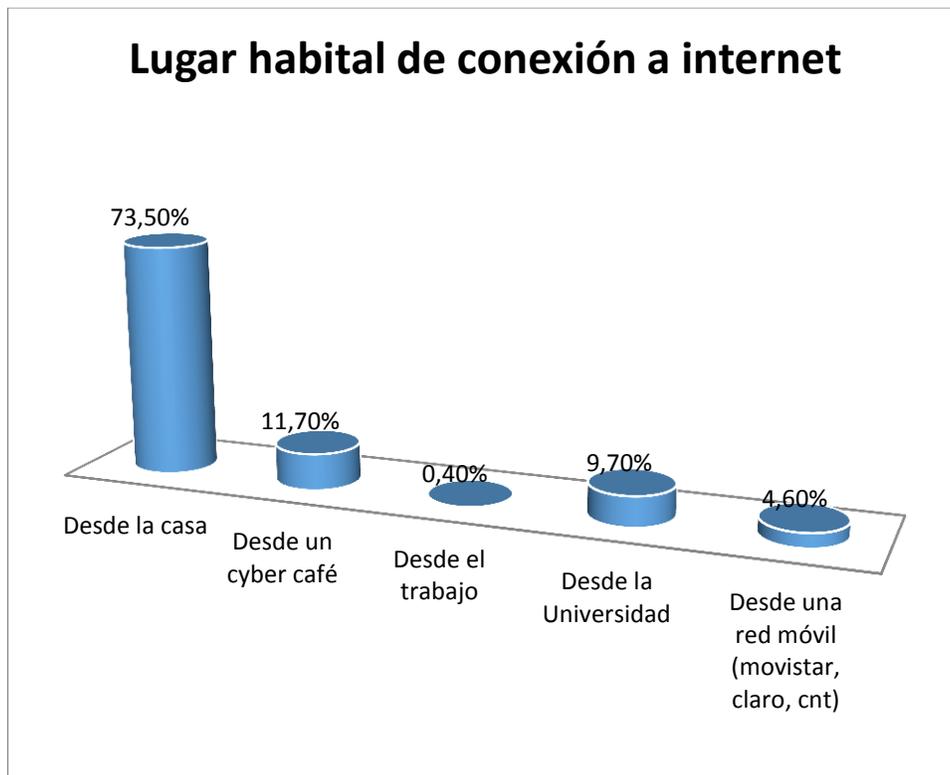


Figura 8: Lugar habitual de conexión a internet
Fuente: Elaboración propia

En lo que se refiere a la cantidad de tiempo en la red, más del 55,3% se conecta como mínimo 7 días. En detalle, el 1,8% se conecta 2 días, el 3,3% 3 días, el 8,0% 4 días, el 17,5% 5 días, 14,2% 6 días, como podemos ver existen porcentajes de conexión de 0,0% 1 día. Se puede concluir que es una población con un nivel alto de uso de Internet. Se estableció una relación entre Ingresos económicos familiares (ing) y días de conexión a internet en la semana (dia_con), obteniendo un valor significativo (R de Pearson 0,322; P= 0,00; P<0,05; C. Determinación = 10,36%; Tau_b = 0,306).

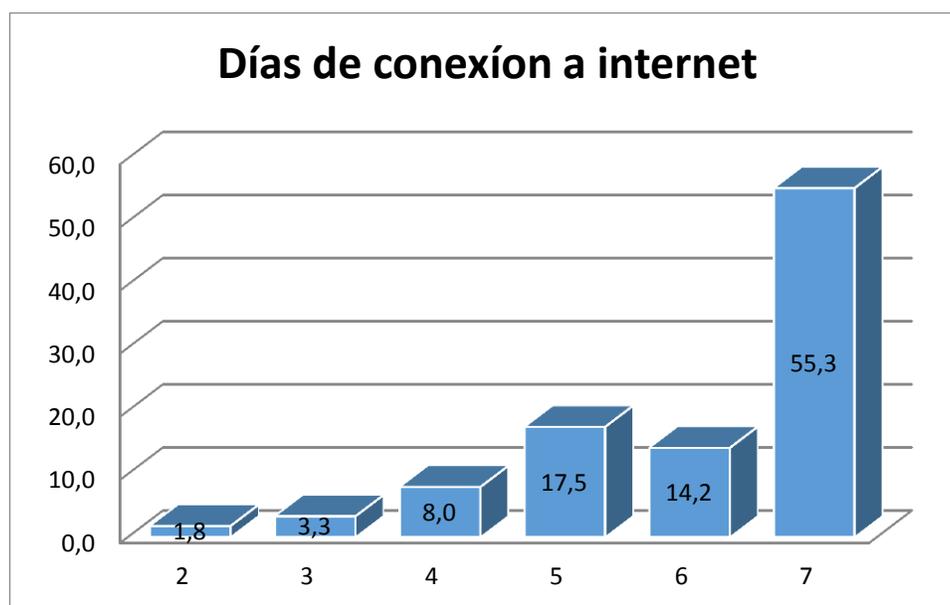


Figura 9: Días de conexión a internet
Figura: Elaboración propia

Los niveles de conocimiento en internet se presentan en la Tabla 8, los cuales se miden en niveles del 1 al 10. Se puede observar que los porcentajes de estudiantes cuyos niveles de conocimiento ubicado en los niveles del 1 y 4 no superan el 7,1%, es decir un porcentaje de 92,9% de estudiantes tienen conocimientos dentro de los niveles de 5 y 10 dando a conocer que tienen conocimientos sólidos.

Se encontró una asociación significativa entre género y nivel de conocimiento de manejo de internet donde un total de una muestra de 452 el porcentaje de 54,0% son hombres y 46,0% mujeres con estos datos ya podemos ver una relación, sin embargo el nivel de significancia se lo comprobó utilizando el estadístico de correlación donde (R de Pearson = -0,148;

$p < 0,002$) refleja una dependencia entre las variables (sex) y (niv_con), calculando el coeficiente de determinación lineal que es de (C. Determinante = 2,19%).

Tabla 8: Conocimiento en Internet

Parámetros	Frecuencia	Porcentaje
2	3	0,7
3	6	1,3
4	23	5,1
5	64	14,2
6	46	10,2
7	95	21,0
8	125	27,7
9	56	12,4
10	34	7,5
Total	452	100,0

Fuente: Datos de la investigación

Elaboración: Propia

Las horas de conexión por día en la Tabla 9, nos presenta del 1 al 15 los niveles de conexión por día, como podemos observar el 97,30% se encuentra en los niveles de conexión de entre el 1 al 10 y el 2,7% del 11 al 15.

Existe una asociación significativa entre la genero (sex) y las horas que se conecta al Internet el estudiante (hor_día). (R. Pearson = 0,101; C. Determinación= 1,4%; $P = 0,003 < 0,05$; Tau_c = 0,100).

Tabla 9: Horas diarias de conexión a internet

Horas	Frecuencia	Porcentaje
1,00 – 5,00	354	78,30
6.00 – 10,00	86	19,00
11.00 – 15,00	12	2,7
Total	452	100%

Fuente: Datos de la investigación

Elaboración: Propia

Abalizada la información sobre, desde hace cuantos años se conecta el estudiante a Internet, la mayor proporción se concentra en los años de 1 - 5 (50,6%), mientras que la menor proporción están alrededor de los 11 a 16 años de experiencia en Internet. Tabla 10. Se encontró una asociación significativa entre edad (eda) y los años de experiencia como internauta (año_exp) (R de Pearson = ,260; C. Determinación = 6,76%; P = 0,000<0,05; Tau_b = 0,166; P= 0,000<0,05; Tau_c = 0,155;).

Tabla 10: Años de experiencia en internet

Años	Frecuencia	Porcentaje
1,00 – 5,00	229	50,60
6,00 – 10,00	208	46,00
11,00 – 16,00	15	3,40
Total	452	100.0

Fuente: Datos de la investigación

Elaboración: Propia

En la tabla 11, podemos observar la intensidad de uso que le dan los estudiantes a los dispositivos móviles dicha tabla refleja que los porcentajes más altos se encuentran en los niveles del 8-10. Así mismo se pudo determinar una asociación significativa entre el uso de dispositivos móviles con los ingresos en detalle Smartphone (R de Pearson = 0,265; P=0,000<0,05; C. Determinate = 7,02%), d_tel_int (R de Pearson = 0,226; P=0,00<0,05; C. Determinante = 5,10%), d_com_por (R de Pearson = 0,133; P=0,005<0,05; C. Determinante = 1,76%), d_tab(R de Pearson = 0,150; P=0,001<0,05; C. Determinante = 2,25%), d_cam_dig (R de Pearson = 0,170; P=0,000<0,05; C. Determinante = 2,89%), d_ipod (R de Pearson= 0,170; P=0,000<0,05; C.Determinante=2,89%), concluyendo que el nivel de uso de dispositivos tiene una dependencia significativa, pero no encontrando esta dependencia ni con la edad ni el género.

Tabla 11: Niveles de uso de dispositivos móviles

Dispositivos							
Niveles de uso	Smartphone	Teléfono móvil con acceso a internet	Teléfono móvil sin acceso a internet	Computador portátil	Table (iPad; Galaxy Tab, Kindle, etc.)	Cámara digital	iPod/MP3 Plyer
Frecuencias							
1	31,6	17,7	33,0	14,2	42,5	27,2	41,6
2	3,1	2,9	5,3	2,7	5,8	6,0	6,6
3	4,2	4,0	6,6	2,7	4,4	9,7	5,5
4	2,9	3,3	5,8	1,5	5,1	9,5	4,9
5	6,6	4,9	7,1	8,0	6,9	7,7	6,6
6	3,1	4,6	4,0	4,9	5,5	5,8	4,6
7	7,3	8,4	5,5	8,0	4,6	6,6	5,8
8	11,9	15,7	5,8	11,5	6,6	9,7	6,2
9	10,2	11,3	7,3	17,3	8,6	6,0	5,8
10	19,0	27,2	19,7	29,4	10,0	11,7	12,4
Total							100,00

Fuente: Datos de la investigación

Elaboración: Propia

4.2. Usos del internet en lo académico y entretenimiento

4.2.1. Usos de índole académico

En este aparatado se muestra los usos de internet para actividades académicas, en lo referente a las asignaturas que se matriculo.

La distribución de ingresos de los estudiantes a la plataforma virtual de la universidad por semana en una escala del 1 al 7 se encuentra con mayor concentración en los porcentajes 35.4% con 160 alumnos y 23.0% con 104 alumnos, de esta forma podemos observar que es habitual ingresar a la plataforma virtual de su universidad, por lo menos de 1 a 3 ingresos semanales.

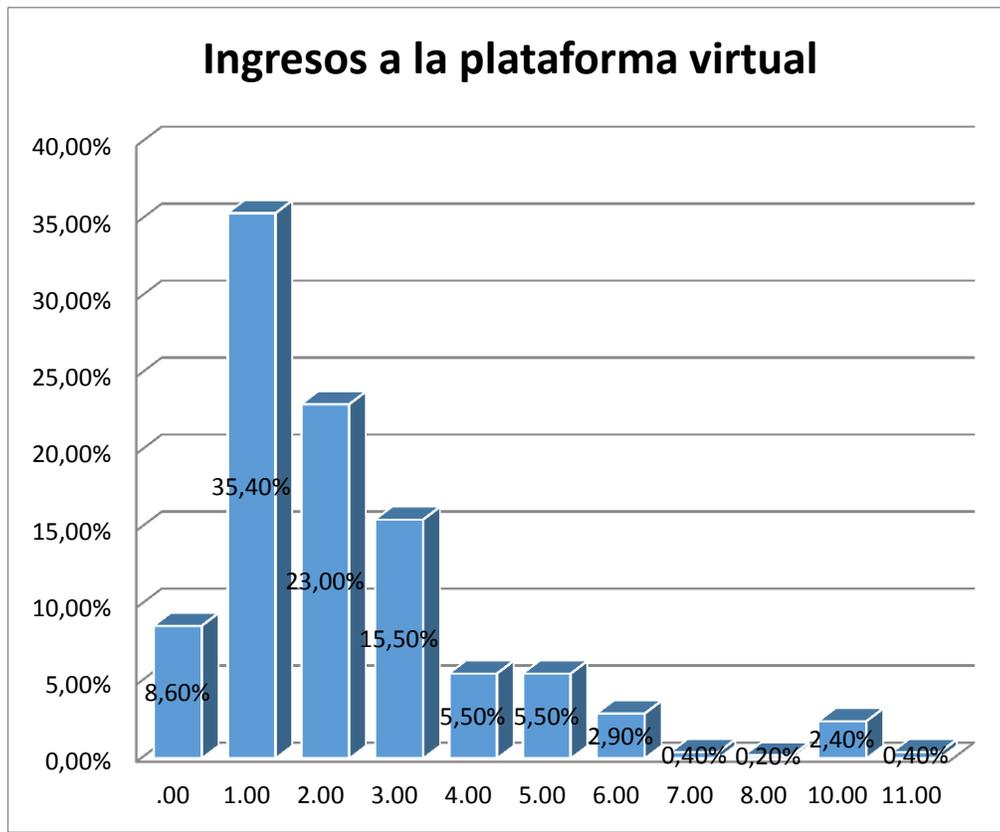


Figura 10: Ingreso a la plataforma virtual
Fuente: Elaboración propia

En la Tabla 12, podemos observar 23,90% realizan de 1 a 4 consultas por mes a sus profesores, seguido por el porcentaje de 21,20% que dice realizar de 9 a 13 consultas mensuales, estos datos determinados por una desviación estándar de 7,74% y una moda de 10%. También podemos ver que un bajo porcentaje del 8,0%, no realiza consultas a sus profesores. En lo referente a consultas a profesores se determinó una variable (a_com_pro) la cual nos presenta una relación significativa con el género (sex) determinado por el valor chi cuadrado ($X^2 = 30,71$; R de Pearson = 0,184; P = 0,00; P = 0,05; C. Determinación = 3,38%).

Tabla 12: Consultas a sus profesores

Consultas	Frecuencia	Porcentaje
0,00	36	8,0
1,00 – 4,00	108	23,90
5,00 – 8,00	81	17,90
9,00 – 13,00	96	21,20
14,00 – 18,00	42	9,30
20,00 – 25,00	74	16,40
27,00 – 30,00	15	3,30
Total	452	100

Fuente: Datos de la investigación

Elaboración: Propia

Consultas que realizan entre compañero de 5 a 8 consultas mensuales con un porcentaje 23,90% con una varianza de 56,6% y una mediana 10%, podemos percatarnos que la interacción entre compañeros sobre temas académicos es muy variada. Como se muestra en la Tabla 13.

Tabla 13: Consultas que realiza a sus compañeros

Consultas	Frecuencia	Porcentaje
0,00	33	7,30
1,00 – 4,00	79	17,50
5,00 – 8,00	108	23,90
9,00 – 12,00	105	23,20
13,00 – 18,00	43	9,50
19,00 – 26,00	65	14,40
27,00 – 36,00	19	4,20
Total	452	100

Fuente: Datos de la investigación

Elaboración: Propia

Las descargas de recursos educativos que realizan los estudiantes se observan en la Tabla 14, en ella las filas indican el nivel de descargas y las columnas los porcentajes con una distribución indistinta en la que la mediana es 4.0 con una varianza de 23.4, donde el número de descargas de recursos educativos (a_rec_edu), se mide en niveles del 1 al 20, obteniendo un mayor porcentaje 14,4%.

Tabla 14: Descargas de recursos educativos

Descargas	Frecuencia	Porcentaje
0,00	80	17,70
1,00 – 5,00	251	55,50
6,00 – 10,00	80	17,70
11,00 – 20,00	41	9,10
Total	452	100

Fuente: Datos de la investigación

Elaboración: Propia

La Tabla 15, indica que alrededor del 55,50% de los estudiantes observan de 1 a 5 videos académicos a través del YouTube cada mes.

Tabla 15: Videos educativos que mira en YouTube cada mes

Búsqueda	Frecuencia	Porcentaje
0,00	80	17,70
1,00 – 5,00	251	55,50
6,00 – 10,00	80	17,70
11,00 – 20,00	41	9,10
Total	452	100

Fuente: Datos de la investigación

Elaboración: Propia

La distribución de la participación en foros virtuales que tiene que realizar los estudiantes se la puede observar en la Figura 11, presentados en escala de 0 al 5 donde 0 significa no acceden a foros y 5 un máximo de foros por mes. Podemos observar que la población estudiantil tiene poca participación en foros.

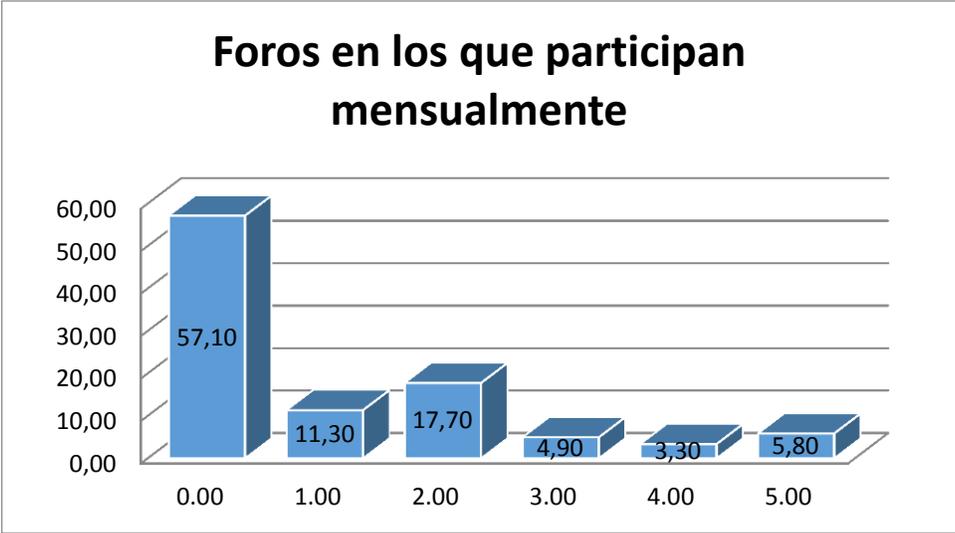


Figura 11: Foros en los que participan mensualmente
Fuente: Elaboración propia

Se puede observar en la Figura 12, 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 22,60% de estudiantes realizan de 3 a 7 posts o tweets al mes.

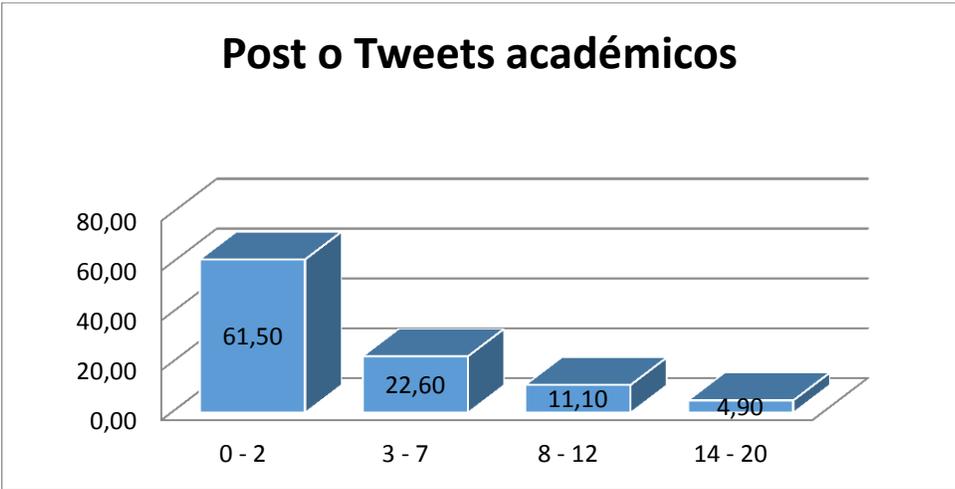


Figura 12: Post o Tweets académicos
Fuente: Elaboración propia

Con respecto del uso de chat para actividades académicas, podemos observar que el promedio de utilización por mes es de 39,40%, con una varianza de 35,81% y una moda 6,0%, mostrado en la Figura 13. Relacionado la edad con las horas que emplean para chatear sobre temas académicos mensualmente se encontró una relación significativa de (χ^2 0,327; P=0,000; P=0,05 C. Determinación=10,6%).

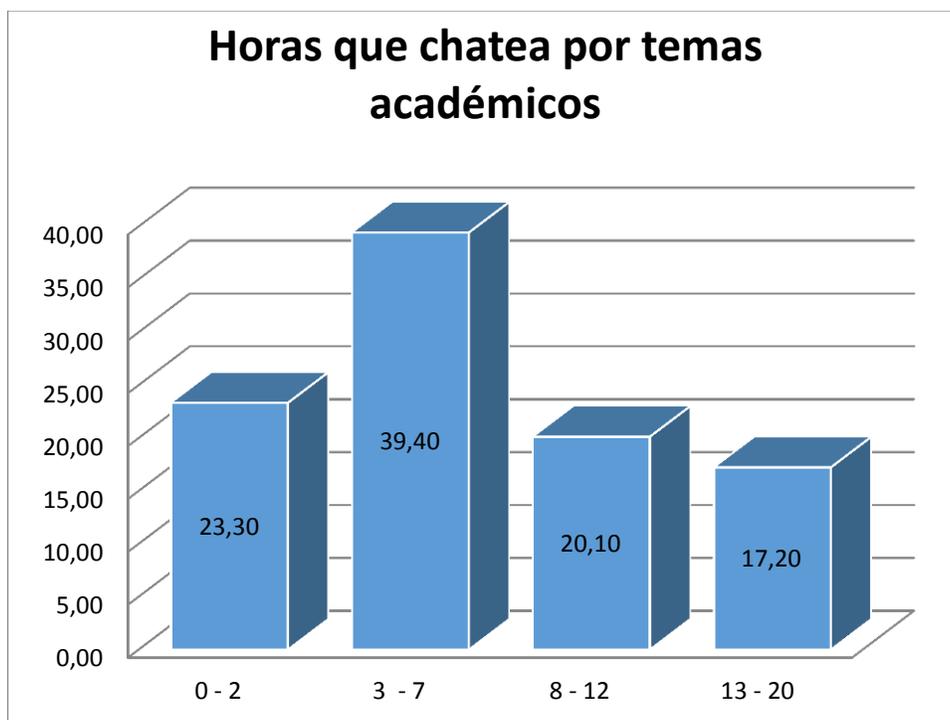


Figura 13: Horas que chatea sobre temas académicos
Fuente: Elaboración propia

En lo referente a información académica que los estudiantes buscan en la web, el mayor número de horas se encuentra en el intervalos de 1 a 10 con un porcentaje de 35,60%, con una desviación estándar del 15,21% y una moda de 20%. Ver Tabla 16. Se determinó una relación significativa entre género y la información académica que buscan los estudiantes mensualmente en la web probado por (R de Pearson = 0,127; P=0,007<0,05; C. Determinante = 1.61%).

Tabla 16: Información académica que busca en internet

Consultas	Frecuencia	Porcentaje
0,00	8	1,80
1,00 – 10,00	161	35,60
11,00 – 20,00	142	31,60
21,00 – 30,00	65	14,40
31,00 – 40,00	32	7,10
41,00 – 60,00	43	9,50
Total	452	100

Fuente: Datos de la investigación

Elaboración: Propia

La mayoría de los estudiantes ingresa a la biblioteca virtual hasta 10 horas mensuales, con un porcentaje de 14,20%, además existe un porcentaje considerable 26,10% de estudiantes que no ingresan a la biblioteca virtual. Figura 14.

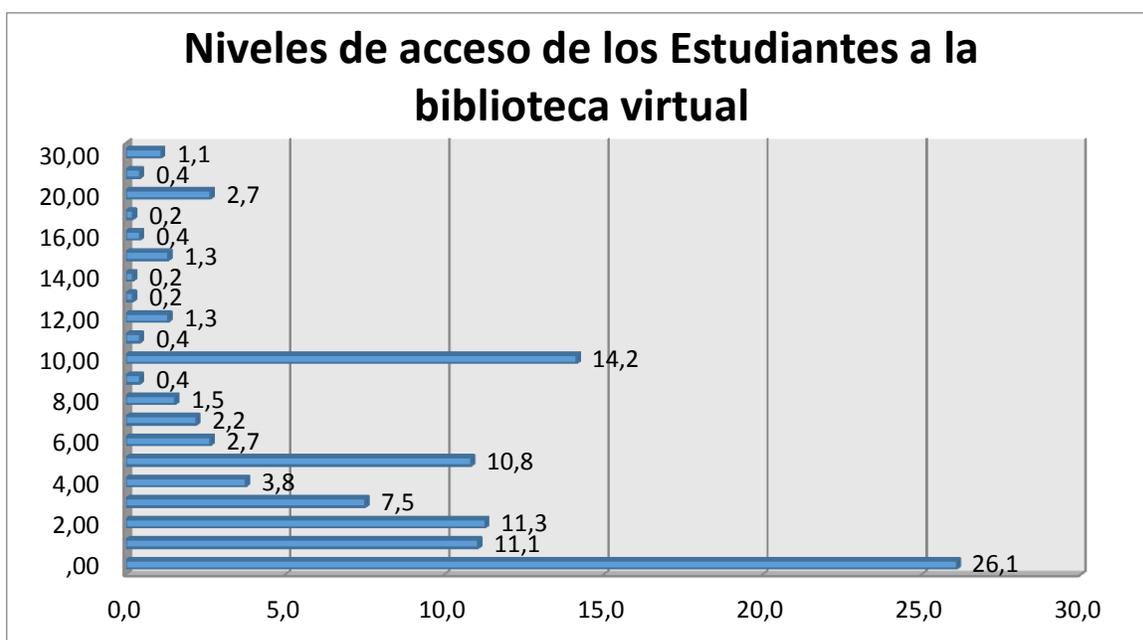


Figura 14: Niveles de acceso de los estudiantes a la biblioteca virtual

Fuente: Elaboración propia

4.2.2. Usos índole no académico o entretenimiento

Al investigar las actividades no académicas en las que participan los estudiantes haciendo uso de internet, podemos ver en que tienen más intensidad de uso y cuales con menores porcentajes para una mejor apreciación se las describe a continuación.

En la Tabla 17, se muestra que la mayor concentración de estudiantes se encuentra en los intervalos de 1 a 5 con un porcentaje de 33,60% y seguidos de los intervalos 6 al 10 con porcentajes iguales, los que permiten constatar las horas semanales que utilizan para chat de entretenimiento o diversión.

Los ingresos (ing) presentan una relación significativa con las horas que chatea por diversión (e_hor_cha), al calcular los valores observados con los esperados Tau-b es de 0,115, con un nivel de significancia de $P= 0,003$. Rechazando la hipótesis nula ya que el valor de $P<0,05$, nos hace comprobar que las variables son independientes entre sí.

Tabla 17: Horas semanales que chatea por diversión

Horas	Frecuencia	Porcentaje
0,00	8	1,80
1,00 – 5,00	152	33,60
6,00 – 10,00	152	33,60
11,00 – 15,00	54	11,90
16,00 – 20,00	47	10,40
21,00 – 25,00	15	3,30
26,00 – 30,00	24	5,4
Total	452	100

Fuente: Datos de la investigación

Elaboración: Propia

Las horas de utilización redes sociales se establecieron, de máximo 40 horas semanales de utilización de redes sociales para diversión (e_red_soc), donde podemos encontrar la mayor concentración en los intervalos de 6-10 horas semanales emplean en redes sociales por diversión con 35,20%, Tabla 18. Los ingresos (ing) determinan una relación significativa con las horas semanales que emplean los estudiantes para chat por diversión (R de Pearson=0,131; P=0,005 P=0,05;) y el nivel de correlación señalado por (Tau_b = 0,113; P=0,003; Tau_c=0,115; P=0,003).

Tabla 18: Horas semanales que utiliza redes sociales

Horas	Frecuencia	Porcentaje
0,00	4	0,90
1,00 – 5,00	118	26,10
6,00 – 10,00	159	35,20
11,00 – 15,00	63	13,90
16,00 – 20,00	51	11,30
21,00 – 25,00	12	2,70
26,00 – 30,00	24	5,30
31,00 – 35,00	21	4,60
Total	452	100

Fuente: Datos de la investigación

Elaboración: Propia

En Tabla 19, se detalla las horas empleadas semanalmente para juegos en línea con un porcentaje del 50,70% no los utiliza. También se encontró una determinación significativa entre las variables género (sex) y horas para juegos en línea (e_jue_lin) con un valor de (R de Pearson=-0,139; P=0,003<0,05) y el coeficiente de determinación lineal de (C. Determinación=1,93%).

Tabla 19: Horas semanales para juegos en línea

Horas	Frecuencia	Porcentaje
0,00	229	50,70
1,00 – 5,00	149	33,0
6,00 – 10,00	45	10,00
11,00 – 15,00	7	1,50
16,00 – 20,00	13	2,80
21,00 – 25,00	9	2,00
Total	452	100

Fuente: Datos de la investigación

Elaboración: Propia

En la Figura 15, determina que las horas que utilizan los estudiantes para realizar descargas de música, videos y programas varía entre los intervalos del 0 al 15, con una desviación estándar de 4,0%.



Figura 15: Horas empleadas a descargas de música, videos y programas

Fuente: Elaboración propia

La Tabla 20, muestra que la mayor proporción de estudiantes 42,70% 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 (R de Pearson = -0,140; P= 0,003<0,05, Tau-b= - 0,122; Tau_c=0,162) con un nivel de significancia P=0,002.

Tabla 20: Videos de entretenimiento que mira en YouTube

Videos	Frecuencia	Porcentaje
0,00	31	6,90
1,00 – 5,00	193	42,70
6,00 – 10,00	126	27,90
11,00 – 15,00	31	6,90
16,00 – 20,00	41	9,10
21,00 – 25,00	10	2,20
26,00 – 30,00	20	4,40
Total	452	100

Fuente: Datos de la investigación

Elaboración: Propia

Los porcentajes más elevados de seguidores y amigos en redes sociales se encuentran en el intervalo de 0, en detalle LinkedIn 90,3%, Twitter 66,6%, porcentajes que afirman que no hacen uso de estas redes sociales, seguido de los intervalos del 1-999, donde se refleja que la red social más utilizada es Facebook con un porcentaje del 83,6%.

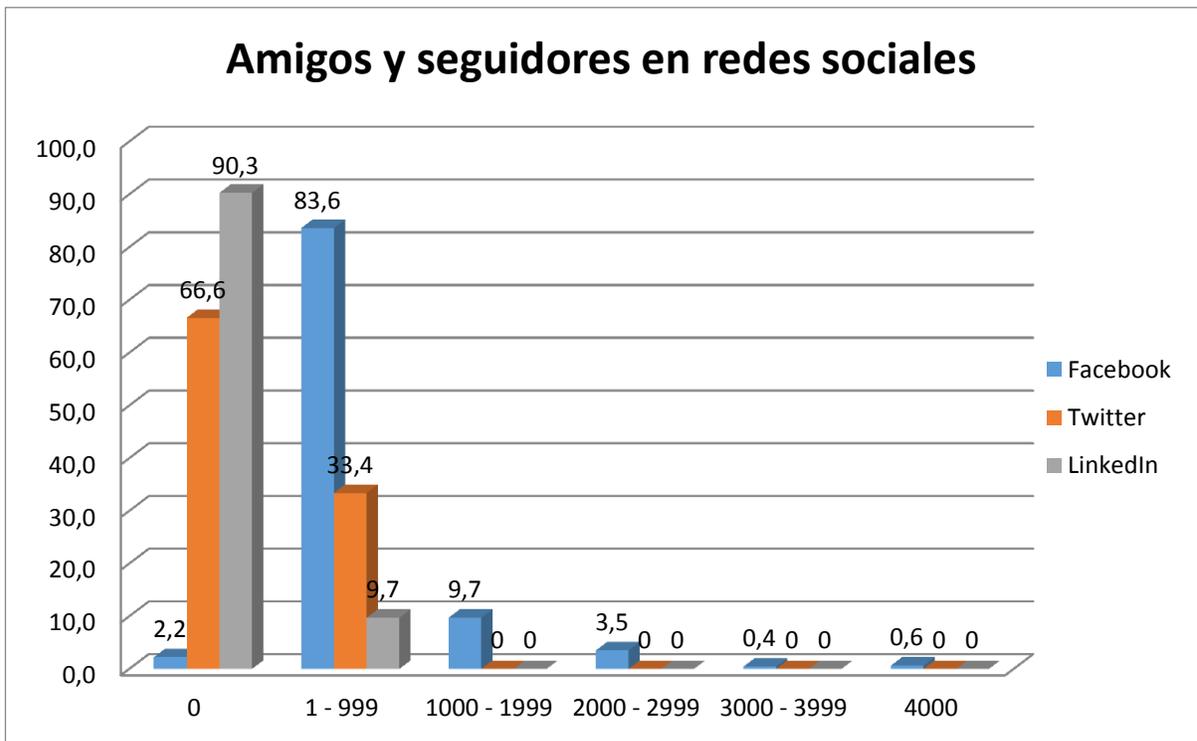


Figura 16: Amigos y seguidores de los estudiantes
Fuente: Elaboración propia

4.2.3. Particularidades rendimiento académicos de los estudiantes

En este apartado se determina el rendimiento académico, para lo cual se hizo uso de la pregunta 17 planteada en la encuesta donde aborda temas tales como: asignaturas matriculadas y asignaturas aprobadas por el estudiante, para construir la variable (rendimiento_ académico), procedimiento que se lo realizó mediante el cálculo de variables, restándole a las asignaturas matriculadas, las asignaturas aprobadas, de los datos obtenidos se los categorizo en niveles (0-1) donde “0” significa las asignaturas aprobadas y “1” las asignaturas reprobadas, este procedimiento se lo realizó teniendo en cuenta que en la universidad donde se está realizando la investigación se lleva el método de matricularse por ciclos donde le recomiendan como máximo coger 9 materias por ciclo. En la Figura 17 podemos ver que de una población de 452 que se encuestó el 77,0% aprueba las asignaturas sin ningún inconveniente y el 23% de alumnado tiende a reprobado de 1 a 3 asignaturas.



Figura 17: Rendimiento académico
Fuente: Elaboración propia

4.3. Perfiles de los estudiantes

En este espacio se trata a detalle las particularidades que presentan los estudiantes, los cuales nos permite agruparlos por categorías para definir los diferentes perfiles como son: uso del internet en actividades académicas (Perfil académico), uso del internet para actividades de entretenimiento (Perfil de entretenimiento), para lo cual aplicamos las diferentes técnicas de minería de datos.

4.3.1. Reducción de variables: Factorización

La técnica de reducción de datos se la aplico con la finalidad de obtener grupos más pequeños que nos puedan aportar datos más representativos de los estudiantes y determinar las preferencias de uso de internet, en la presente investigación se la aplico en el uso de internet en lo académico y para entretenimiento, donde pudimos concluir que la aplicación de esta técnica no es factible debido a que no se cumplían las principales condiciones como, la determinante de la matriz de correlaciones no es cercana a uno, la medida muestral KMO dio como resultado un valor bajo, el valor de la varianza total no llego al valor mínimo de 70%, como podemos ver en (Anexos 5 y 6 y 7).

4.3.2. Clusterización

Luego de descartar la técnica de factorización, la misma que no cumplió con lo requerido para ser aplicable en la presente investigación, se procedió aplicar la técnica de clusterización, con la finalidad de agrupar a los estudiantes en base a grupos diferenciados en el uso de internet tanto en lo académico y entretenimiento descartando todas las formas posibles de agrupación. Para el perfil académico, concerniente a la pregunta 10, con la finalidad de escoger la mejor agrupación se desarrolló cuatro distribuciones en grupos de 2, 3, 4 y 5, (Anexos 8, 9, 10 y 11) el mismo procedimiento se aplicó para el perfil de entretenimiento de la pregunta 11, en el cual se formaron grupos de 2, 3, 4 y 5. (Anexos 12, 13, 14 y 15).

4.3.3. Representación del perfil académico

El perfil académico se lo elaboro en base a las variables que nos permiten determinar las características de los estudiantes a la hora de hacer uso del internet en actividades académicas, estas variables se las tomo de la pregunta 10 (Anexo 1), en las que solo intervinieron 5 las cuales son: consultas que hacen los estudiantes a los profesores y compañeros, videos académicos que mira en YouTube, horas que chatea sobre temas académicos, horas que busca información académica en internet, todas estas actividades están medidas de manera mensual.

Se creó los conglomerados o clúster utilizando el algoritmo de k-medias que nos permite procesar un número ilimitado de casos, pero sólo permite utilizar un método de aglomeración y requiere que se proponga previamente el número de conglomerados que se desea obtener. (Spss, 2010). Para verificar la clasificación obtenida, los grupos fueron creados en dos etapas, calculando primero los centroides de las clasificaciones en base a una submuestra, y luego con base a los centroides se generaron los grupos.

(Díaz de Rada, 2002), recomienda el mayor porcentaje de clasificación, para la selección del clúster más adecuado, esto se realizó aplicando la técnica de discriminación que tiene como objetivo discriminar los resultados obtenidos en el análisis clúster para comprobar el nivel de exactitud para cada grupo. Los resultados de precisión de las clasificaciones de los grupos se los muestra en la Tabla 21, considerando el porcentaje de clasificación más alto

99,30% y tomando en cuenta el principio de interpretabilidad, que sugiere el autor, la clasificación seleccionada para el perfil académico es la del grupo de 2.

Tabla 21. Nivel de exactitud para grupos de actividades académicas

Clasificación en grupos	Porcentaje de exactitud
Grupo de 2	99,30%
Grupo de 3	97,60%
Grupo de 4	96,70%
Grupo de 5	97,80%

Fuente: Datos de la investigación

Elaboración: Propia

En este caso se tomó la clasificación de dos grupos porque es la muestra más representativa y la que más cercana a uno, permitiéndonos hacer una mejor interpretación de los datos, como se lo muestra en la Tabla 22, en esta tabla se puede verificar los porcentajes de discriminación de las variables constatando que el grupo dos tiene niveles más altos de uso de internet en actividades académicas por lo que se los auto denominó como regular al grupo 1 y dedicado al grupo 2.

Tabla 22: Discriminación de variables académicas

Variables	Función discriminante	
	Grupo 1	Grupo 2
Consultas a profesores	8,80	11,77
Consultas a compañeros	8,73	12,77
Videos académicos mira en YouTube	6,32	8,70
Chat sobre temas académicos	6,59	8,19
Información académica que busca en internet	12,01	40,19

Fuente: Datos de la investigación

Elaboración: Propia

Teniendo a consideración las variables descritas en la Tabla 22, se ha clasificado en dos grupos los mismos que permiten una mayor diferenciación e sus centroides.

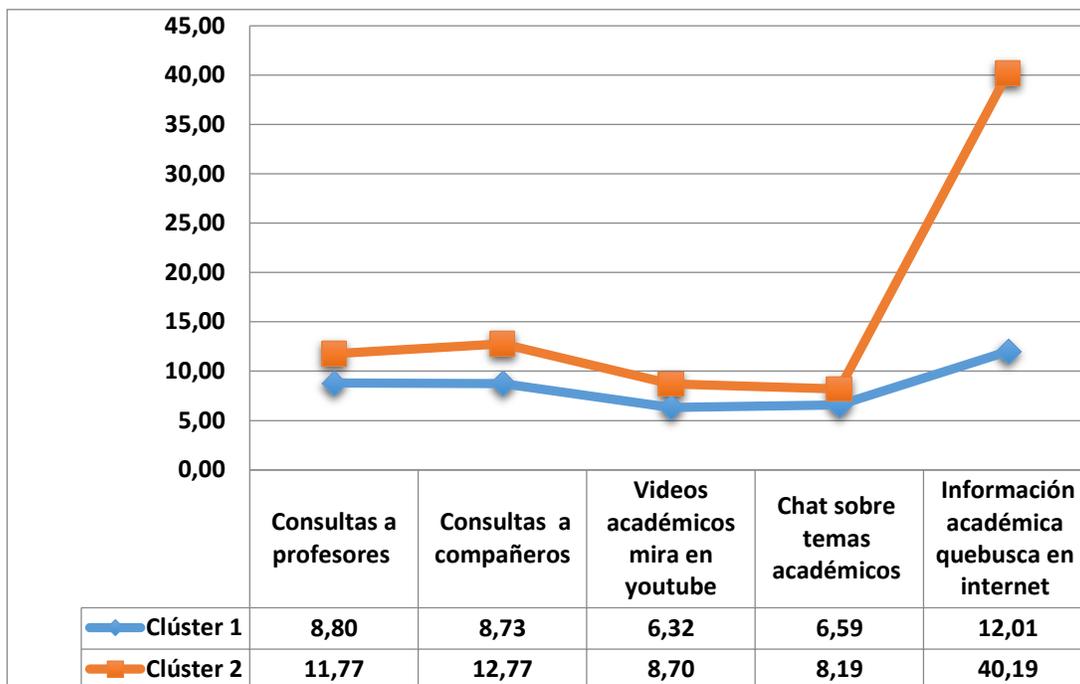


Figura 18: Centroides del perfil académico
Fuente: Elaboración propia

El grupo 1 nos da la interpretación de ser un grupo **Regular**, en comparación con el grupo dos, esto debido a los niveles de actividad que presentan porcentajes menores. Los estudiantes de este grupo tiene características poco activas, en las diferentes actividades tales como hacer consultas a profesores, a compañeros, chatean menos horas sobre temas académicos, miran menos videos en internet y buscan menos información académica mensualmente, se conforma por un porcentaje que alcanza el 28,50% de los estudiantes, están entre una edad de 20 años, en su mayoría son hombres con el 58,20% y mujeres que alcanza el 41,8%, los estudiantes de este grupo en su mayoría se conecta desde un cyber café y tiene un nivel bajo en experiencia en uso de internet, y poca conexión a la semana.

El grupo dos se conforma del 71,50% de estudiantes y sus principales características son reflejadas en los porcentajes de actividad (Figura 18), todas estas actividades por encima del grupo uno por ello lo denominamos grupo **Dedicado**. Habitualmente este grupo ingresa a internet desde la casa, tiene un nivel de ingresos mayores, así como más años de experiencia en el manejo de internet, este grupo en su mayoría son mujeres alcanzando un porcentaje de 56,6% con una edad promedio de 24 años y con el 43,4% de hombres.

Los estudiantes que pertenecen al grupo de dedicados son los que más hacen uso de internet para actividades que tiene que ver con rendimiento académico, esto se debe a que pertenecen a familias con mayores ingresos, los cuales están por encima de los \$600,00 con un porcentaje de 33,2%, a diferencia de los del grupo regular que en su mayoría tiene ingresos familiares de menos de \$350,00 que alcanzan el 34,7%, además el grupo activo tiene internet en casa lo cual se le hace más fácil acceder al mismo viendo reflejado esto con el 73,5%, a diferencia del grupo pasivo que solo hacen uso de internet ingresando desde un cyber café, haciendo las diferencias más visibles, de uso de internet entre los grupos de estudiantes, así mismo el grupo activo hace más uso de las redes móviles como (movistar, claro, cnt), este grupo pertenece a carreras como ingeniería en turismo con 15,3% , licenciatura en contabilidad y computación 5,1%, licenciatura en diseño gráfico 15,0% Ingeniería en Mantenimiento Automotriz con el 8,8%.

4.3.4. Representación del perfil de entretenimiento

Para obtener el perfil de entretenimiento se siguió el mismo procedimiento que para obtener el perfil académico como está planteado en el apartado anterior, en el cual se seleccionó las variables con medias más altas de la pregunta 11 (Anexo 1), en las cuales intervienen las variables como: cuantas horas chatea por diversión, cuantas horas utiliza redes sociales, cuantas horas utiliza juegos en línea, horas que descarga música, videos y programas, videos de entretenimiento que mira en YouTube, variables que fueron medidas semanalmente.

De los resultados obtenidos del análisis clúster creado con el algoritmo k-medias, se aplicó análisis discriminante, que consiste en poner como variable de agrupación el resultado de cada uno de los grupos y como variables independientes a cada una de las variables que están siendo analizadas, para conocer los usos de internet para entretenimiento. Determinando de esta manera el porcentaje de precisión de las clasificaciones como se muestra en la Tabla 23, se puede constatar que la clasificación de dos y tres grupos, tienen el mismo porcentaje de discriminación con un valor del 97,6%, sin embargo considerando que lo que se busca es establecer diferencias claras entre los usuarios y tomando en cuenta como quedaron en la forma gráfica de cada uno de los grupos, se ha creído conveniente escoger la clasificación de tres grupos.

Tabla 23. Nivel de exactitud para grupos de entretenimiento

Agrupaciones	Porcentaje de distribución
Grupo de 2	97,6%
Grupo de 3	97,6%
Grupo de 4	95,4%
Grupo de 5	96,9%

Fuente: Datos de la investigación

Elaboración: Propia

En este caso se ha seleccionado la muestra de tres grupos que es la que nos permite hacer la mejor representación de los grupos autodenominados: grupo uno intensivo, grupo dos activo y grupo tres pasivo.

Tabla 24: Discriminación de variables de entretenimiento

Variables	Función discriminante		
	Grupo 1	Grupo 2	Grupo 3
Chat por diversión	20,47	12,06	6,21
Redes sociales	27,92	11,37	7,61
Juegos en línea	4,71	7,13	1,00
Descargas de música, videos y programas	6,52	6,84	3,37
Videos de entretenimiento que mira en YouTube	11,86	16,75	4,15

Fuente: Datos de la investigación

Elaboración: Propia

El uso de internet que hacen los estudiantes para entretenimiento de alguna forma tiene relación con el desempeño académico de los estudiantes, permitiéndoles desarrollar destrezas de navegación, búsqueda de información y acceso a contenidos, pero el uso inadecuado en la utilización de internet puede traer graves consecuencias como fomentar adicciones ocasionando un bajo rendimiento. Para verificar lo anterior dicho hemos clasificado en esta etapa una descripción que diferencia a los estudiantes en función de sus

características más importantes; en la Figura 19, se pueden observar los centroides de los clúster para cada uno de los factores utilizados.



Figura 19: Centroides de perfil de entretenimiento
Fuente: Elaboración propia

El grupo 1 conformado por el 18,40% de la población estudiantil, se caracteriza por tener menos actividad en el uso de internet para entretenimiento, esta agrupado casi parcialmente entre mujeres y hombres, los cuales alcanzan porcentajes de 50,4% y 49,6% respectivamente, a este grupo se lo ha denominado **Pasivo**. Los valores más altos que ascienden en la utilización de internet para entretenimiento es la interacción a través de chat, redes sociales y ver videos en YouTube, sin embargo estos porcentajes está muy por debajo de los grupos dos y tres, con lo referente a juegos en línea tiene una actividad de nula (cero), igualmente las actividades de descarga de música, videos y programas están dentro de los factores con valores más bajos. Estos niveles de utilización de la tecnología permiten delinear un perfil de usuario para quien el entretenimiento es poco importante o no

tiene las posibilidades tecnológicas o de tiempo, o sus prioridades con el uso de internet son otras diferentes a las de entretenimiento. La edad promedio de este grupo es de 22 años, con porcentajes de población significativamente menores que los otros grupos.

El grupo 2 está conformado por el 21,50% de los estudiantes, se mantiene en un nivel intermedio en actividades de uso de internet para entretenimiento, este grupo está conformado con un porcentaje de 61,9% de hombres y 38,1% de mujeres, tiene una edad promedio de 20 años, la principal característica de este grupo es hacer notar que tiene mayor porcentaje de actividad, en cuanto a lo que tiene que ver con mirar videos de entretenimiento en youtube por semana, realizando la actividad por encima de los otros grupos. A pesar de no ser tan altos los porcentajes de los otros factores tales como: chat por diversión, redes sociales, juegos en línea y descargas de música, videos y programas hace notar que son usuarios que si acceden a estos programas. Denominado como un grupo **Activo**.

El grupo 3 conformado por el 60,10% de los estudiantes, sus integrantes se caracterizan por tener mayor actividad en uso de redes sociales, chat por diversión, con menores porcentajes pero por encima de la escala del grupo uno y dos, también realiza las actividades tales como: juegos en línea, descargas de música, videos, programas, y mirar videos de entretenimiento en YouTube, por tal motivo que se lo ha denominado como grupo **Intensivo**, es su mayoría por hombres con el 59,0% con una media de edad entre los 18 años, son estudiantes que están empezando su carrera universitaria.

4.4. Fase de evaluación e interpretación de datos

En esta fase de minería de datos se trata de la comprobación de las hipótesis planteadas para este proyecto de investigación donde se aplica técnicas de chi cuadrado y modelos de regresión logística que nos permiten verificar si las hipótesis propuestas son prudentes y significativas.

4.4.1. El nivel de ingresos determina como se utiliza internet para el aprendizaje

En este apartado se comprobó la hipótesis uno: el nivel de ingresos determina como se utiliza el internet para el aprendizaje: proceso que se lo enfatizo mediante una variable independiente y una variable dependiente:

Variable independiente: Niveles de ingreso los cuales se los describió en el apartado anterior los que constan de 5 niveles en detalle; [ing=1,00] = 350 dólares, [ing=2,00] =600 dólares, [ing=3,00]= 1.000 dólares, [ing=4,00] = 1.500 dólares, [ing=5,00] = más de 1.500 dólares, determinada como una variable categórica, donde la mayoría de estudiantes se concentra en los niveles 1 y 2 con el 34,70% y 33,20% respectivamente.

Variable dependiente: Resultados del clúster de “perfil académico”, que consta de dos categorías: Regular = 1, con un porcentaje de 28,50% y Dedicado = 2, con un porcentaje de 71,50%.

Para comprobar la relación entre las dos variables se utilizó la prueba de Chi Cuadrado, los resultados obtenidos son: ($X^2 = 1,288$; $P= 0,863$; $P=0,05$) con 4 grados de libertad donde podemos ver que el valor de significancia es mayor que $P= 0,05$, concluyendo que no hay relación significativa entre las dos variables Tabla 7.

Se realizó la comprobación del modelo mediante la regresión logística binaria debido a que el resultado del “perfil académico” tomado como variable dependiente asume dos categorías: “Regular” y “Dedicado”, para esto se recategorizó en una variable de tipo Dummy o binaria y se toma como referencia los ingresos familiares.

La prueba Omnibus, que nos permite hacer la comprobación de que al menos uno de los coeficientes del modelo sea diferente de cero, donde X^2 nos da un resultado de ($X^2 = 1,322$; $P=0,858$), obteniendo un valor mayor a ($P>0,05$), lo que determina que el modelo no es estable o robusto.

La bondad del ajuste del modelo a través de Hosmer y Lemeshow obtiene un valor ($X^2 = 0,000$; $P= 1,000$; $P>0,05$), por lo tanto no es significativo, debido a que la prueba que se le realiza al modelo implica que se ajuste a los datos, el estadístico R cuadrado de Nagelkerke tiene un valor de 0,004 (0,4%) de la varianza que explica el modelo. Constatando que los valores observados son iguale a los predichos. (Anexo 18).

La verificación de los coeficientes de las variables independiente se lo realizo mediante la prueba de Wald que se detalla en la Tabla 25, sin embargo podemos observar que las variables muestran valores no significativos ($P > 0,05$). Refiriéndonos a los porcentajes obtenidos de la hipótesis, se rechaza la hipótesis nula y se acepta la hipótesis alternativa, concluyendo que no existe una relación significativa entre el uso de internet en actividades académicas y el nivel de ingresos.

Tabla 25: Coeficientes del modelo de regresión logística de los usos de internet en actividades académicas e ingresos

Variables en la ecuación									
	B	Error estándar	Wald	gl	Sig.	Exp(B)	95% C.I. para EXP(B)		
							Inferior	Superior	
Paso 1 ^a									
Ing			1,280	4	,865				
ing(1)	,100	,510	,038	1	,845	1,105	,407	3,002	
ing(2)	,165	,510	,105	1	,746	1,179	,434	3,207	
ing(3)	-,043	,537	,006	1	,937	,958	,335	2,745	
ing(4)	-,272	,624	,190	1	,663	,762	,224	2,591	
Constante	-,981	,479	4,198	1	,040	,375			

a. Variables especificadas en el paso 1: ing.

Fuente: Datos de la investigación

Elaboración: Propia

4.4.2. El nivel de ingresos determina como se utiliza internet para entretenimiento

Con respecto a la segunda hipótesis referente a si el nivel de ingresos determina como se utiliza internet para entretenimiento, partiendo del apartado anterior donde se describió los ingresos económicos familiares, los mayores porcentajes se encuentran en los niveles 1 y 2 con el 34,70% y 33,20% respectivamente, para la comprobación de esta hipótesis el atributo ingresos económicos también es considerado como independiente, mientras que para describir el uso de internet en actividades de entretenimiento se considera el perfil de entretenimiento, el mismo que consta de tres categorías etiquetadas como: Intensivo, Activo y Pasivo, el cual se lo toma como dependiente.

Para comprobar la relación entre las dos variables se utilizó la prueba de Chi Cuadrado, los resultados obtenidos son: ($X^2 = 21,059$; $P= 0,007$; $P>0,05$) con 8 grados de libertad donde podemos ver que el valor de significancia es menores que $P= 0,05$, concluyendo que los ingresos si determinan el uso de internet en entretenimiento Tabla 6.

Comprobando el modelo mediante regresión logística multinomial se presenta la varianza que explica si el modelo es estable o robusto representado por el valor de Nagelkerke = 0,051, es decir el 5,10%. La prueba de Wald en la Tabla 26, nos muestra que en el modelo realizado para comprobar la hipótesis en el nivel de Ingresos 1 para el grupo Pasivo y el Ingreso 1 para el grupo Activo son significativos ($P<0,05$).

El primer modelo OR de pertenencia al perfil de entretenimiento Pasivo con respecto al perfil Intensivo es 0,156 veces menor cuando un estudiante pertenece al nivel de Ingreso 1 respecto al nivel Ingreso 5 ($OR=0,153$ (IC 95% 0,051 – 0,478) $P=0,001$).

En el segundo modelo, el OR de pertenencia al perfil de entretenimiento Activo con respecto al perfil de entretenimiento Pasivo es de 0,200 veces menor cuando un estudiante pertenece al nivel de Ingreso 1 respecto al nivel de Ingresos 5 ($OR=0,200$ (IC 95% 0,062 – 0,648) $P=0,007$) (Anexo 19).

En resumen de las pruebas expuestas arrojan un análisis del modelo que demuestran que los niveles de ingresos de las familias de los estudiantes no representan incidencias mayores o significativas en ninguno de los grupos que forman parte del perfil de entretenimiento, debido a que en los modelos expuestos solo en Ingreso 1 determina el uso del internet para los grupos de entretenimiento Pasivo y Activo con respecto al grupo Intensivo. Con estos resultados se considera pertinente rechazar la hipótesis nula, y dar por aceptada la hipótesis alternativa que plantea que el nivel de ingresos no determina como se utiliza internet para entretenimiento.

Tabla 26: Coeficientes del modelo de regresión logística de los usos de internet en actividades de entretenimiento e ingresos

Estimaciones de parámetro									
Número de caso de clúster ^a		B	Error estándar	Wald	gl	Sig.	Exp(B)	95% de intervalo de confianza para Exp(B)	
								Límite inferior	Límite superior
Pasivo	Interceptación	,405	,527	,592	1	,442			
	[ing=1,00]	-1,859	,572	10,569	1	,001	,156	,051	,478
	[ing=2,00]	-1,708	,573	8,881	1	,003	,181	,059	,557
	[ing=3,00]	-1,504	,597	6,351	1	,012	,222	,069	,716
	[ing=4,00]	-1,232	,695	3,142	1	,076	,292	,075	1,139
	[ing=5,00]	0 ^b	.	.	.	0	.	.	.
Activo	Interceptación	,154	,556	,077	1	,782			
	[ing=1,00]	-1,608	,599	7,206	1	,007	,200	,062	,648
	[ing=2,00]	-1,179	,592	3,967	1	,046	,307	,096	,981
	[ing=3,00]	-1,142	,618	3,413	1	,065	,319	,095	1,072
	[ing=4,00]	-,362	,670	,292	1	,589	,696	,187	2,589
	[ing=5,00]	0 ^b	.	.	.	0	.	.	.
a. La categoría de referencia es: Intensivo									
b. Este parámetro está establecido en cero porque es redundante.									

Fuente: Datos de la investigación

Elaboración: Propia

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

En este apartado se verifica la hipótesis H_3 la misma que hace énfasis en si el uso de la tecnología en el aprendizaje incide en el rendimiento académico.

Partiendo del rendimiento _ académico, variable que se construyó en el apartado 4.2.3, a partir de las asignaturas matriculados y asignaturas aprobadas, obteniendo una variable de tipo categórica de dos niveles en detalle: 0 = “Aprobado”, 1 = “Reprobado”.

Para la obtención del modelo se aplicó Regresión Logística Binaria, debido a las características de las variables. La variable “rendimiento _ académico” considerada como

dependiente y el resultado del perfil académico “Clúster _ académico” como la variable independiente.

La prueba de bondad de ajuste de Ómnibus para la hipótesis H_3 nos da como resultado valor de Chi cuadrado ($X^2 = 1,453$; $P=0,228$) con 1 grado de libertad que no es significativo $P>0,05$. Al comprobar la bondad de ajuste mediante el resumen del modelo el Logaritmo de la verosimilitud -2 da un valor 481,268. El valor R cuadrado de Nagelkerke en este modelo alcanza un valor de varianza 0,005 (5%) (Anexo 20).

En la comprobación de los coeficientes de la variables independientes se empleó la prueba de test de Wald como resultado obtuvimos valores no significativos ($p>0,05$). Tabla 27.

Tabla 27: Coeficientes del modelo de regresión logística binaria (H_3)

Variables en la ecuación							
		B	Error estándar	Wald	gl	Sig.	Exp(B)
Paso 1 ^a	Cluster_academico	,294	,242	1,478	1	,224	1,342
	Constante	-1,616	,338	22,818	1	,000	,199

a. Variables especificadas en el paso 1: Cluster_academico.

Fuente: Datos de la investigación

Elaboración: Propia

Podemos comprobar que el valor de Odds Ratio (OR) no representa relaciones significativas. Como resultado se rechaza la hipótesis H_3 y se concluye que no existe incidencia en el uso del internet en actividades académicas sobre el rendimiento académico.

Para tener una mejor veracidad del de los resultados obtenidos en el procedimiento anterior se finalizó aplicando la prueba de Chi Cuadrado para medir la diferencia entre los valores esperados y los valores observados los valores obtenidos de dicha prueba son ($X^2 = 1,484$; $P = 0,223$; $P> 0,05$) verificando que no existe incidencia significativa entre las variables dependientes e independientes Tabla 7.

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

De la misma manera también se verifico la hipótesis H_4 la misma que hace referencia en si el uso de la tecnología en el entretenimiento incide en el rendimiento académico.

Partiendo del rendimiento_académico, variable que se construyó en el apartado 4.2.3, a partir de las asignaturas matriculados y asignaturas aprobadas, obteniendo una variable de tipo categórica de dos niveles en detalle: 0 = “Aprobado”, 1 = “Reprobado”.

Para la obtención del modelo se aplicó Regresión Logística Binaria, debido a las características de las variables. La variable “rendimiento _ académico” considerada como dependiente y el resultado del Perfil de Entretenimiento “Cluster_entretenimiento” como la variable independiente.

La prueba de bondad de ajuste de Ómnibus para la hipótesis H_4 nos da como resultado valor de Chi cuadrado ($X^2 = 0,274$; $P=0,601$) con 1 grado de libertad que no es significativo $P>0,05$. Al comprobar la bondad de ajuste mediante el resumen del modelo el Logaritmo de la verosimilitud - 2 da un valor 482,447. El valor R cuadrado de Nagelkerke en este modelo alcanza un valor de varianza 0,005 (5%) (Anexo 21).

En la comprobación de los coeficientes de la variables independientes se empleó la prueba de test de Wald como resultado obtuvimos valores no significativos ($p>0,05$). Tabla 28.

Tabla 28: Coeficientes del modelo de regresión logística binaria (H_4)

Variables en la ecuación							
		B	Error estándar	Wald	gl	Sig.	Exp(B)
Paso 1 ^a	Cluster_entretenimiento	-,075	,142	,276	1	,599	,928
	Constante	-1,053	,358	8,642	1	,003	,349

a. Variables especificadas en el paso 1: Cluster_entretenimiento.

Fuente: Datos de la investigación

Elaboración: Propia

Podemos comprobar que el valor de Odds Ratio (OR) no representa relaciones significativas. Como resultado se acepta la hipótesis nula que concluye que no existe

incidencia en el uso del internet en actividades de entretenimiento sobre el rendimiento académico.

Para finalizar se aplicó la prueba de Chi Cuadrado para medir la diferencia entre los valores esperados y los valores observados los valores obtenidos de dicha prueba son ($X^2 = 1,233$; $P = 0,540$; $P > 0,05$) verificando que no existe incidencia significativa entre las variables dependientes e independientes Tabla 4.

4.5. Análisis y discusión de los resultados

En esta apartado hace una discusión de los hallazgos encontrados en la investigación, de cada una de las hipótesis planteadas en este proyecto de investigación, con la finalidad de determinar la incidencia de los ingresos familiares en los usos de internet en actividades académicas y de entretenimiento, además del rendimiento académico de los alumnos.

Partiendo de la población universitaria en la misma que está conformada por 452 estudiantes encuestados, los que tienen una edad promedio de 18 a 36 años, con porcentajes mayores entre las edades de 18 a 20 años, de los cuales el 54% son hombres y el 46% son mujeres.

Los ingresos familiares de los estudiantes de la UTN son bajos se encuentran en los niveles del 1 al 2 de 350\$ a 600\$ dólares con un porcentaje de 67,90%, pero podemos concluir que aunque los ingresos familiares no son altos, no es un impedimento para optar por una formación de nivel superior. De este grupo podemos recalcar que el nivel que pasan conectados por semana es alto donde más del 50% se conecta como mínimo los 7 días a la semana, siendo el lugar más habitual de conexión a la web desde la casa con 73,50%.

4.5.1. Determinante del nivel de ingresos en el uso de internet en actividades académicas

Los resultados obtenidos en la Hipótesis 1 después de la aplicación de técnicas minería de datos contrastan la siguiente información: Se pudo comprobar que existen datos estadísticos no significativos, debido a que las constantes no llegan a explicar la pertenencia de la variable independiente. El no encontrar, que los niveles de ingresos familiares no tengan

incidencia o no determinen el uso de Internet para actividades académicas no se la ve como una causante para que los estudiantes no obtengan conocimientos sólidos en el manejo de internet.

Al aplicar la prueba de Chi Cuadrado en el aparato anterior, los resultados que se obtuvieron son: ($X^2 = 1,288$; $P = 0,863$; $P = 0,05$) con 4 grados de libertad donde podemos ver que el valor de significancia es mayor que $P = 0,05$, concluyendo que no hay relación significativa entre las dos variables Tabla 7.

Esta información la podemos afianzar con el estudio realizado por (Ayala, 2007) donde pudo encontrar que no existen relaciones significativas entre el nivel de ingresos con el logro académico, comprobado a través de la hipótesis que se plantea a mayores ingresos económicos, mayores logros académicos.

Podemos recalcar que estos resultados no concuerdan con estudios realizados por (Torres & Infante, 2011) donde afirman que los ingresos sí inciden en el uso de internet en actividades académicas, a mayor ingreso del núcleo familiar del estudiante, mayor son las posibilidades de contar con un computador y acceso a internet.

4.5.2. Determinante del nivel de ingresos en el uso de internet en actividades de entretenimiento

Las técnicas aplicadas para dar solución a la Hipótesis 2, nos arrojaron resultados que nos demuestran que el modelo no es robusto a pesar de encontrar valores significativos $P < 0,05$ en nivel de ingresos 1 equivalente a 350 dólares para el perfil de entretenimiento en los grupos Pasivo y Activo con respecto a Intensivo. Demostrando que no existe una relación significativa entre el nivel de ingresos económicos y el uso de Internet en actividades de entretenimiento.

De los grupos conformados en el perfil académico el grupo 3 denominado como Intensivo con el 60,10% de los estudiantes tienen mayores porcentajes de uso de redes sociales y chat por diversión. Según los datos obtenidos en esta investigación los niveles de ingresos de las familias de los estudiantes de la UTN son bajos, pero esto no les impide que hagan usos elevados de internet para entretenimiento donde con mayores porcentajes realizan actividades como chat por diversión con el 33,60% 5 a 11 horas semanales y el 35,20%

permanece 6 a 10 horas semanales haciendo uso de redes sociales, (Ayala, 2007) en su estudio nos muestra que el uso de internet para entretenimiento no muestra mayores diferencias entre grupos con ingresos económicos altos o bajos, pero tiende a ser más fuerte en grupos de menores ingresos.

4.5.3. Incidencia del uso de internet en actividades académicas sobre rendimiento académico

La Hipótesis 3 hace referencia al uso de internet para actividades académicas y su influencia en el rendimiento académico en esta investigación no demostró una relación significativa, demostrando que el uso de internet no define el rendimiento académico del estudiante de la UTN.

Aplicando la prueba de Chi Cuadrado para medir la diferencia entre los valores esperados y los valores observados los valores obtenidos de dicha prueba son ($X^2 = 1,484$; $P = 0,223$; $P > 0,05$) verificando que no existe incidencia significativa entre las variables dependientes e independientes, por lo que se rechaza la Hipótesis nula y se acepta la Hipotesis alternativa que el uso de internet para actividades académicas no incide en el rendimiento académico Tabla 7.

4.5.4. Incidencia del uso de internet en actividades de entretenimiento sobre rendimiento académico

Para comprobar la Hipótesis 4 uso de internet para actividades de entretenimiento y su influencia en el rendimiento académico se aplicó el modelo predictivo de regresión logística binomial, podemos constatar que no varían los resultados en las pruebas que conforman el modelo, verificando que no cumplen las condiciones que permite aceptar el modelo y por lo tanto se da paso a captar la hipótesis alternativa.

Para finalizar se aplicó la prueba de Chi Cuadrado para medir la diferencia entre los valores esperados y los valores observados los valores obtenidos de dicha prueba son ($X^2 = 1,233$; $P = 0,540$; $P > 0,05$) verificando que no existe incidencia significativa entre las variables dependientes e independientes Tabla 4.

CONCLUSIONES Y RECOMENDACIONES

Conclusiones

En este apartado se recopila todas las conclusiones que han surgido a través del desarrollo de este proyecto de investigación.

De los ingresos familiares establecidos en cinco niveles, los datos obtenidos de los estudiantes de la Universidad Técnica del Norte, se encuentran en los niveles 1 y 2: es decir 350\$ y 600\$ mensuales, dando a conocer que los ingresos familiares son bajos, pero no es una limitante para optar por una educación de nivel superior.

Se comprobó que no existe una relación significativa entre ingresos y usos de internet en actividades académicas y de entretenimiento los modelos aplicados con regresión logística no se ajustaron a las pruebas.

Al describir las principales técnicas empleadas en dar solución al proyecto de investigación se pudo constatar que las más factibles por los datos obtenidos: fueron la clusterización, junto algoritmo de K-means, nos permite obtener las características de los estudiantes agrupados según sus preferencias.

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 ($P=0,223$; $P>0,05\%$).

Finalmente ya finalizada la investigación y comprobadas las hipótesis planteadas en el proyecto de investigación, podemos concluir que el proyecto: la educación virtual en el Ecuador, representa una valiosa fuente de información y comunicación sobre el uso de la tecnología en las universidades ecuatorianas.

Recomendaciones

En este apartado se recopila todas las recomendaciones que han surgido a través del desarrollo de este proyecto de investigación para futuras investigaciones.

Los ingresos económicos familiares de los estudiantes de la Universidad Técnica del Norte fueron bajos y los niveles de uso de internet en actividades de entretenimiento elevados, actividades que en su mayoría las realizan desde casa, para esto se debería concientizar a los estudiantes de los esfuerzos de sus familias por darles accesibilidad a una herramienta como es internet para que hagan uso de ella adecuadamente y sacarle el mejor potencial en beneficio académico.

En esta investigación se encontró que el uso de internet en lo académico como entretenimiento no inciden sobre el rendimiento académico, se debería ampliar el análisis aplicando en la encuestas variables que permitan medir más detalladamente en que están empleando los estudiantes el uso de internet.

Al finalizar la investigación se hace necesario recomendar para investigaciones futuras la dependencia entre variables, con la finalidad de obtener resultados más veraces que aporten a los objetivos de la investigación.

Se recomienda utilizar el software SPSS para la realización de minería de datos, pues es una herramienta muy completa que permite aplicar un sinnúmero de técnicas de minería de datos para proyectos de investigación de tipo estadístico.

BIBLIOGRAFÍA

- Ayala, C. (2007). *Relación entre el uso de Internet y el Logro Académico*.
- Cardenas, M. (2013). *Clustering: Clasificación no Supervisada Gráficas estadística y minería de datos con python*.
- Carrasco, J., & Martínez, J. (2011). *Komputer Sapiens*. Obtenido de Reconocimiento de Patrones.
- Castaño, J. (2011). *El uso de internet para la interacción en el aprendizaje: un análisis de la eficacia y la igualdad en el Sistema Universitario Catalán*. Obtenido de http://www.tesisenred.net/bitstream/handle/10803/52561/Tesis_Jonatan_Casta%C3%B1o.pdf?sequence=1.
- 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*.
- Castaño, C. (2008). La segunda brecha digital y las mujeres jóvenes. *Quaderns de la Mediterrània*, 11, 218-224.
- Castaño, C., Martín, J., & Martínez, J. (2011). *La brecha digital de género en España y Europa: medición con indicadores compuestos*. oi:10.5477/cis/reis.136.127.
- Castellanos, M. (2009). *Formula para calcular poblaciones finitas*.
- De la Fuente, S. (2011). *Análisis de correspondencias simples y múltiples*. Obtenido de <http://www.fuenterrebollo.com/Economicas/ECONOMETRIA/REDUCIR-DIMENSION/CORRESPONDENCIAS/correspondencias.pdf>
- Dialnet. (2010). *Dialnet*. Obtenido de <https://dialnet.unirioja.es/descarga/articulo/4786672.pdf>
- Díaz de Rada, V. (2002). *Técnicas de análisis multivariante para investigación social y comercial*. (R.-M. S.A, Ed.). Madrid, España.
- DiMaggio, P., & Harggitai, E. (2001). *From the "Digital Divide" to 'Digital Inequality': Studying Internet Use As Penetration Increases*. New Jersey: Princeton University. Obtenido de <https://www.princeton.edu/~artspol/workpap/WP15 - DiMaggio%2BHargittai.pdf>

- DiMaggio, P., Hargittai, E., Celeste, C., & Shafer, S. (2004). From Unequal Access to Differentiated Use: A Literature Review and Agenda for Research on Digital Inequality. *En K. (Russell S. F. Neckerman (Ed.), Social inequality*, (pp. 355–400). New York: Citeseer.
- DiMaggio, P., Hargittai, E., Neuman, R., & Robinson, J. (2001). Social Implications of the Internet. *Annals of Physics*, 54, 258.
- Gallardo, E. (2012). Hablemos de estudiantes digitales y no de nativos digitales. *UT. Revista de Ciències de l'Educació*, Pag. 7-21.
- Graham, R. (30 de Septiembre de 2008). *The Stylisation of Internet Life?: Predictors of Internet Leisure Patterns Using Digital Inequality and Status Group Perspectives*. Obtenido de <<http://www.socresonline.org.uk/13/5/5.html>> doi:10.5153/sro.1804
- Gutierrez, L., & Gamboa, L. (2010). Determinants of ICT Usage among Low-Income Groups in Colombia, Mexico, and Peru. *The Information Society* 26: 346–363, DOI: 10.1080/01972243.2010.511559.
- Hargittai, E. (2010). Digital Na(t)ives? Variation in Internet Skills and Uses among Members of the “Net Generation”. *Sociological Inquiry*, Vol. 80, No. 1, 92–113 DOI: 10.1111/j.1475-682X.2009.00317.x.
- Hargittai, E., & Shafer, S. (2006). Differences in Actual and Perceived Online Skills: The Role of Gender. *Social Science Quarterly*, (87)2.
- Hasperué, W. (2012). *Extracción de conocimiento en grandes bases de datos utilizando estrategias adaptativas*. Obtenido de Universidad Nacional de la Plata: http://postgrado.info.unlp.edu.ar/Carreras/Doctorado/Tesis/Hasperue_Waldo.pdf
- Hernández, J., Ramírez, J., & Ferri, C. (2007). *Introducción a la Minería de Datos*. Mexico: Ed. Pearson Educación.
- INEC. (2013). *Reporte Anual de Estadísticas de la Tecnologías de la Información y Comunicaciones (TIC'S)*.
- Jain, A. (2009). *Data clustering: 50 years beyond K-means*. ScienceDirect.
- Jansen, J., & Fellow, S. (2010). *Use of the internet in higher-income households*. Obtenido de <http://pewinternet.org/Reports/2010/Better-off-households.aspx>

- Moreno, G. (2007). *Técnicas más usadas en Minería de Datos*. Obtenido de <https://gamoreno.wordpress.com/2007/10/03/tecnicas-mas-usadas-en-la-mineria-de-datos/>
- Pando Fernández, & San Martín Fernández. (2004). *Regresión Logística Multinomial*.
- Pérez, C., & Santín, D. (2007). *Minería de Datos Técnicas y Herramientas*. Madrid: Ed. Thomson ISBN:978-9732-482-2.
- Porcel, E., Dopaso, G., & López, M. (2008). *Modelos predictivos y técnicas de minería de datos para la identificación de factores asociados al rendimiento académico de alumnos universitarios*.
- Serrano, A., & Martínez, E. (2003). *La Brecha Digital: Mitos y Realidades*. México: (UABC, Ed.).
- SITEAL. (2012). *Sistema de Información de Tendencias Educativas en America Latina*. Obtenido de http://www.siteal.iipe-oei.org/datos_destacados/372/la-brecha-digital-en-america-latina
- Taylor, W., Zhu, G., Dekkers, J., & Marshall, S. (2003). Socio-Economic Factors Affecting Home Internet Usage Patterns in Central Queensland. *Informing Science Journal*, 6, 233–246.
- Tello, E. (Enero de 2008). *Las tecnologías de la información y comunicaciones (TIC) y la brecha digital: su impacto en la sociedad de México*.
- Torres, J., & Infante, A. (2011). *Desigualdad digital en la universidad: usos de Internet en Ecuador*.
- 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*.
- UIT. (2012). *Unión Internacional de Telecomunicaciones*. Obtenido de Recuperado de: https://www.itu.int/dms_pub/itu-d/opb/ind/D-IND-ICTOI-2012-SUM-PDF-S.pdf
- UNESCO. (2005). *Hacia las sociedades del conocimiento*. Obtenido de <http://www.unesco.org/publications>

Van Dijk, J. (2006). Digital divide research, achievements and shortcomings. *Poetics* 34 (2006) 221–235, doi:10.1016/j.poetic.2006.05.004.

ANEXOS

Anexo 1: Encuesta aplicada al estudiante

La educación virtual en Ecuador: Proyecto de investigación

Estimado estudiante, solicitamos su colaboración contestando esta encuesta, la que permitirá desarrollar una investigación para conocer el uso de internet en las universidades del Ecuador.

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

2. Responda la siguiente pregunta	
¿Que carrera estudia?	

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

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

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

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

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

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

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

10. En lo referente a las asignaturas en las que está matriculado	
¿Cuántas veces a la semana ingresa a la plataforma virtual de su universidad?	()
¿Aproximadamente cuántas consultas les hace a sus profesores cada mes?	()
¿Aproximadamente cuántas consultas realiza a sus compañeros por mes?	()
¿Aproximadamente cuántos recursos educativos descarga de la plataforma virtual cada mes?	()
¿Aproximadamente cuántos videos académicos mira en youtube cada mes?	()
¿Aproximadamente en cuántos foros virtuales participa cada mes?	()
¿Aproximadamente cuántos post o tweets sobre temas académicos realiza en las redes sociales por mes?	()
¿Aproximadamente cuántas horas chatea sobre temas académicos cada mes?	()
¿Aproximadamente cuántas horas busca información académica en internet cada mes?	()
¿Aproximadamente cuántas horas utiliza la biblioteca virtual de la universidad cada mes?	()

11. En lo referente al entretenimiento y diversión en internet	
¿ Aproximadamente cuántas horas a la semana chatea por diversión?	()
¿Aproximadamente cuántas horas a la semana utiliza redes sociales?	()
¿Aproximadamente cuántas horas a la semana utiliza juegos en línea?	()
¿Aproximadamente cuántas horas a la semana descarga música, videos y programas?	()
¿Aproximadamente cuántos videos para entretenimiento mira en youtube cada semana?	()

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

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

14. ¿Cuál es su nivel de uso de los siguientes dispositivos? (1 significa no usar y 10 significa utilizar al máximo)	1	2	3	4	5	6	7	8	9	10
Smartphone con cámara fotográfica y acceso a internet	()	()	()	()	()	()	()	()	()	()
Teléfono móvil con acceso a internet	()	()	()	()	()	()	()	()	()	()
Teléfono móvil sin acceso a internet	()	()	()	()	()	()	()	()	()	()
Computador portátil	()	()	()	()	()	()	()	()	()	()
Tablet (iPad; Galaxy Tab, Kindle, etc.)	()	()	()	()	()	()	()	()	()	()
Cámara digital	()	()	()	()	()	()	()	()	()	()
iPod / MP3 Player	()	()	()	()	()	()	()	()	()	()

15. De 1 a 10 valore los siguientes aspectos (1 significa no estar de acuerdo y 10 estar completamente de acuerdo)										
	1	2	3	4	5	6	7	8	9	10
Internet le permite elaborar los trabajos más rápido y con menos esfuerzo	()	()	()	()	()	()	()	()	()	()
Usted confía en la información de internet para realizar sus tareas	()	()	()	()	()	()	()	()	()	()
Internet le permite prescindir de la Biblioteca	()	()	()	()	()	()	()	()	()	()
Internet facilita el proceso de aprendizaje	()	()	()	()	()	()	()	()	()	()
Internet le permite mejorar sus calificaciones	()	()	()	()	()	()	()	()	()	()
Usted presenta trabajos académicos copiados desde Internet	()	()	()	()	()	()	()	()	()	()

16. Responda las siguientes preguntas referentes a sus profesores. (Se recomienda evaluar de forma general a todos sus profesores)	SI	NO	A VECES
Su profesor ingresa a la plataforma virtual	()	()	()
Contesta sus consultas por correo electrónico	()	()	()
Chatea con usted eventualmente sobre aspectos académicos	()	()	()
Su profesor comenta en redes sociales sobre temas académicos	()	()	()
Le sube materiales digitales a la plataforma virtual	()	()	()
Le recomienda recursos digitales de la biblioteca virtual	()	()	()
Le recomienda videos sobre temas académicos	()	()	()
Le plantea cuestionarios o evaluaciones en la plataforma virtual	()	()	()
Le plantea foros virtuales	()	()	()
Su profesor tiene una página web, blog o perfil de facebook	()	()	()
Su profesor tiene cuenta de twitter	()	()	()

17. Responda las siguientes preguntas:	
En el semestre anterior, ¿en cuántas asignaturas se matriculó?	()
En el semestre anterior ¿cuántas asignaturas aprobó?	()

Anexo 2: Guía de entrevista para un(a) directivo de la institución educativa

1. Instalaciones físicas

- ¿Cuenta con salas de cómputo para los estudiantes?
- ¿Cuenta con Internet inalámbrico para los estudiantes?
- ¿Ancho de banda total?
- ¿Número de estudiantes?
- ¿Tiene software para la gestión académica y administrativa?
Desarrollo propio _____
Sistema comercial _____
- Cuenta con un campus virtual (entorno virtual de aprendizaje)?
Desarrollo propio: _____
Sistema comercial: _____ Nombre: _____
- No tiene _____
- ¿Los estudiantes se pueden matricular en línea?
- ¿Los estudiantes pueden consultar sus calificaciones en línea?
- ¿Los estudiantes y profesores tienen acceso a una biblioteca virtual?
- ¿Los profesores suben las calificaciones a Internet?
- ¿Qué porcentaje de profesores cuenta con un computador personal?

2. Políticas de uso de tecnología

- ¿Los docentes utilizan obligatoriamente elementos tecnológicos en sus actividades de enseñanza?
Si _____ NO _____
- ¿Cuenta la institución con un plan de formación de profesores en el uso de tecnologías para la educación?
- ¿Aproximadamente qué porcentaje de sus profesores está capacitado en temas tecnológicos?
- ¿Aproximadamente qué porcentaje de sus profesores está capacitado en temas pedagógicos?
- ¿Tienen algún tipo de incentivo para los profesores que utilizan las tic en sus actividades académicas?

Anexo 3: Carta para el levantamiento de información



Oficio Nro.391-DGITT-UTPL
Loja, 16 de Septiembre de 2014

Doctor
Miguel Naranjo Toro
Rector de la Universidad Técnica del Norte
En su despacho.-

De mi consideración:

A través del presente le hago extensivo un saludo cordial y los mayores deseos de éxito en sus actividades académicas.

A más de saludarlo, solicito su colaboración con un proyecto de investigación que busca determinar cómo se utiliza la tecnología en las actividades académicas en las universidades ecuatorianas, este proyecto está a cargo de 35 estudiantes y en el caso de la institución que usted dirige, el (la) señor(a) Lady Susana Jimenez Abad con cédula 1104542699 es quien tiene la labor de aplicar 500 encuestas a estudiantes y una entrevista a un directivo de la institución, cuyo modelo se adjunta.

La tarea que realizará el (la) señor(a) Jimenez Abad le servirá como trabajo de graduación y los resultados de la investigación a más de ser publicados, servirán como línea de base para la educación superior del país. Por tal motivo, y conoedores de su espíritu de colaboración con la investigación y desarrollo del país, le solicito se permita al (la) mencionado(a) señor(a) aplicar las encuestas en los predios universitarios.

Solicito confirmar su autorización al correo electrónico lsjimenezx@utpl.edu.ec o al correo jctorres@utpl.edu.ec

Por la atención que se dé al presente, le antelo mi agradecimiento.

Atentamente



Juan Pablo Suárez Chacón Ph. D.
DIRECTOR GENERAL DE INVESTIGACIÓN
UNIVERSIDAD TÉCNICA PARTICULAR DE LOJA



San Cayotano Alto s/n
Loja-Ecuador
Tel: (593-7) 2570275
Fax: (593-7) 2584893
Apartado Postal: 11-01-608
info@utpl.edu.ec
www.utpl.edu.ec

Anexo 4: Carta de aceptación a levantamiento de información



RECTORADO

UNIVERSIDAD TÉCNICA DEL NORTE

Resolución No. 001- 073 CEAACES - 2013 - 13

Ibarra, 13 de octubre del 2014

Oficio 851-R

PhD
Juan Pablo Suárez Chacón
**DIRECTOR GENERAL DE INVESTIGACIÓN
UNIVERSIDAD TÉCNICA PARTICULAR DE LOJA**

De mi consideración:

En relación a su atenta comunicación 391-DGITT-UTPL, al respecto autorizo a la **señorita Lady Susana Jiménez Abad**; a fin de que aplique las 500 encuestas a los señores alumnos y la entrevista a un directivo de la Universidad Técnica del Norte; como trabajo de graduación de la mencionada profesional, y los resultados de la investigación que a más de ser publicados, servirán como línea de base de la Educación Superior del País.

Con sentimientos de consideración y aprecio.

Cordialmente,

CIENCIA Y TÉCNICA AL SERVICIO DEL PUEBLO


Doctor Miguel Naranjo Toro
RECTOR



cvr

Av. 17 de Julio s - 21 y José María Córdova
Ciudadela Universitaria Barrio El Olivo
Teléfono: (593) 6 2997800 Casilla 199
Fax: Ext. 7011
Email: rectorado@utn.edu.ec
www.utn.edu.ec
Ibarra - Ecuador

Anexo 5: Análisis factorial para usos de internet en actividades académicas

Matriz de correlaciones ^a						
a. Determinante = ,673						
Prueba de KMO y Bartlett						
Medida Kaiser-Meyer-Olkin de adecuación de muestreo					,663	
Prueba de esfericidad de Bartlett	Aprox. Chi-cuadrado				177,504	
	gl				10	
	Sig.				,000	
Varianza total explicada						
Componente	Autovalores iniciales			Sumas de extracción de cargas al cuadrado		
	Total	% de varianza	% acumulado	Total	% de varianza	% acumulado
1	1,835	36,691	36,691	1,835	36,691	36,691
2	,965	19,300	55,991			
3	,884	17,672	73,663			
4	,715	14,299	87,962			
5	,602	12,038	100,000			
Método de extracción: análisis de componentes principales.						

Anexo 6: Análisis factorial usos de internet en actividades de entretenimiento

Matriz de correlaciones ^a						
a. Determinante = ,348						
Prueba de KMO y Bartlett						
Medida Kaiser-Meyer-Olkin de adecuación de muestreo					,713	
Prueba de esfericidad de Bartlett	Aprox. Chi-cuadrado				473,558	
	gl				10	
	Sig.				,000	
Varianza total explicada						
Componente	Autovalores iniciales			Sumas de extracción de cargas al cuadrado		
	Total	% de varianza	% acumulado	Total	% de varianza	% acumulado
1	2,361	47,229	47,229	2,361	47,229	47,229
2	,854	17,089	64,318			
3	,812	16,230	80,548			
4	,641	12,826	93,374			
5	,331	6,626	100,000			
Método de extracción: análisis de componentes principales.						

Anexo 7: Análisis factorial nivel de uso de dispositivos

Matriz de correlaciones ^a						
a. Determinante = ,566						
Prueba de KMO y Bartlett						
Medida Kaiser-Meyer-Olkin de adecuación de muestreo					,664	
Prueba de esfericidad de Bartlett	Aprox. Chi-cuadrado				255,306	
	gl				6	
	Sig.				,000	
Varianza total explicada						
Componente	Autovalores iniciales			Sumas de extracción de cargas al cuadrado		
	Total	% de varianza	% acumulado	Total	% de varianza	% acumulado
1	1,942	48,547	48,547	1,942	48,547	48,547
2	,830	20,759	69,306			
3	,774	19,356	88,662			
4	,454	11,338	100,000			
Método de extracción: análisis de componentes principales.						

Anexo 8: Uso de internet en actividades académicas clúster de dos grupos

Centros de clústeres finales		
	Clúster	
	1	2
¿Aproximadamente cuántas consultas les hace a sus profesores cada mes?	8.80	11.77
¿Aproximadamente cuántas consultas realiza a sus compañeros por mes?	8.73	12.77
¿Aproximadamente cuántos videos académicos mira en youtube cada mes?	6.32	8.70
¿Aproximadamente cuántas horas chatea sobre temas académicos cada mes?	6.59	8.19
¿Aproximadamente cuántas horas busca información académica en internet cada mes?	12.01	40.19

ANOVA						
	Clúster		Error		F	Sig.
	Media cuadrática	gl	Media cuadrática	gl		
¿Aproximadamente cuántas consultas les hace a sus profesores cada mes?	810.726	1	58.210	450	13.928	.000
¿Aproximadamente cuántas consultas realiza a sus compañeros por mes?	1504.501	1	53.349	450	28.201	.000
¿Aproximadamente cuántos videos académicos mira en youtube cada mes?	522.992	1	27.011	450	19.362	.000
¿Aproximadamente cuántas horas chatea sobre temas académicos cada mes?	234.434	1	35.364	450	6.629	.010
¿Aproximadamente cuántas horas busca información académica en internet cada mes?	73203.407	1	69.310	450	1056.173	.000

Las pruebas F sólo se deben utilizar con fines descriptivos porque los clústeres se han elegido para maximizar las diferencias entre los casos de distintos clústeres. Los niveles de significación observados no están corregidos para esto y, por lo tanto, no se pueden interpretar como pruebas de la hipótesis de que los medias de clúster son iguales.

Número de casos en cada clúster		
Clúster	1	323.000
	2	129.000
Válido		452.000
Perdidos		.000

Anexo 9: Uso de internet en actividades académicas clúster de tres grupos

Centros de clústeres finales			
	Clúster		
	1	2	3
¿Aproximadamente cuántas consultas les hace a sus profesores cada mes?	6.32	11.74	13.09
¿Aproximadamente cuántas consultas realiza a sus compañeros por mes?	7.04	12.42	12.47
¿Aproximadamente cuántos videos académicos mira en youtube cada mes?	5.19	9.07	8.44
¿Aproximadamente cuántas horas chatea sobre temas académicos cada mes?	4.80	6.62	10.18
¿Aproximadamente cuántas horas busca información académica en internet cada mes?	8.51	48.16	22.66

ANOVA						
	Clúster		Error		F	Sig.
	Media cuadrática	gl	Media cuadrática	gl		
¿Aproximadamente cuántas consultas les hace a sus profesores cada mes?	2321.858	2	49.803	449	46.621	.000
¿Aproximadamente cuántas consultas realiza a sus compañeros por mes?	1652.121	2	49.459	449	33.404	.000
¿Aproximadamente cuántos videos académicos mira en youtube cada mes?	677.881	2	25.217	449	26.882	.000
¿Aproximadamente cuántas horas chatea sobre temas académicos cada mes?	1355.926	2	29.925	449	45.311	.000
¿Aproximadamente cuántas horas busca información académica en internet cada mes?	43734.304	2	37.693	449	1160.265	.000
Las pruebas F sólo se deben utilizar con fines descriptivos porque los clústeres se han elegido para maximizar las diferencias entre los casos de distintos clústeres. Los niveles de significación observados no están corregidos para esto y, por lo tanto, no se pueden interpretar como pruebas de la hipótesis de que las medias de clúster son iguales.						

Número de casos en cada clúster		
Clúster	1	215.000
	2	73.000
	3	164.000
Válido		452.000
Perdidos		.000

Anexo 10: Uso de internet en actividades académicas clúster de cuatro grupos

Centros de clústeres finales						
	Clúster					
	1	2	3	4		
¿Aproximadamente cuántas consultas les hace a sus profesores cada mes?	18.75	11.47	6.31	6.10		
¿Aproximadamente cuántas consultas realiza a sus compañeros por mes?	15.91	12.29	7.18	7.49		
¿Aproximadamente cuántos videos académicos mira en youtube cada mes?	9.27	8.95	4.98	6.92		
¿Aproximadamente cuántas horas chatea sobre temas académicos cada mes?	10.53	6.48	4.50	8.37		
¿Aproximadamente cuántas horas busca información académica en internet cada mes?	21.22	48.10	6.48	21.48		
ANOVA						
	Clúster		Error		F	Sig.
	Media cuadrática	gl	Media cuadrática	gl		
¿Aproximadamente cuántas consultas les hace a sus profesores cada mes?	3769.024	3	35.040	448	107.563	.000
¿Aproximadamente cuántas consultas realiza a sus compañeros por mes?	1904.389	3	44.193	448	43.093	.000
¿Aproximadamente cuántos videos académicos mira en youtube cada mes?	485.243	3	25.050	448	19.371	.000
¿Aproximadamente cuántas horas chatea sobre temas académicos cada mes?	819.651	3	30.556	448	26.825	.000
¿Aproximadamente cuántas horas busca información académica en internet cada mes?	29757.176	3	33.753	448	881.612	.000
Las pruebas F sólo se deben utilizar con fines descriptivos porque los clústeres se han elegido para maximizar las diferencias entre los casos de distintos clústeres. Los niveles de significación observados no están corregidos para esto y, por lo tanto, no se pueden interpretar como pruebas de la hipótesis de que los medias de clúster son iguales.						

Número de casos en cada clúster		
Clúster	1	93.000
	2	73.000
	3	171.000
	4	115.000
Válido		452.000
Perdidos		.000

Anexo 11: Uso de internet en actividades académicas clúster de cinco grupos

Centros de clústeres finales					
	Clúster				
	1	2	3	4	5
¿Aproximadamente cuántas consultas les hace a sus profesores cada mes?	4.69	18.63	10.49	16.81	6.05
¿Aproximadamente cuántas consultas realiza a sus compañeros por mes?	6.30	11.80	9.90	21.17	7.91
¿Aproximadamente cuántos videos académicos mira en youtube cada mes?	4.82	7.30	9.25	9.68	7.17
¿Aproximadamente cuántas horas chatea sobre temas académicos cada mes?	4.36	7.86	5.56	11.21	8.78
¿Aproximadamente cuántas horas busca información académica en internet cada mes?	6.45	13.42	49.80	31.06	21.67

ANOVA						
	Clúster		Error		F	Sig.
	Media cuadrática	gl	Media cuadrática	gl		
¿Aproximadamente cuántas consultas les hace a sus profesores cada mes?	3557.641	4	28.578	447	124.487	.000
¿Aproximadamente cuántas consultas realiza a sus compañeros por mes?	2342.994	4	36.106	447	64.891	.000
¿Aproximadamente cuántos videos académicos mira en youtube cada mes?	346.828	4	25.259	447	13.731	.000
¿Aproximadamente cuántas horas chatea sobre temas académicos cada mes?	625.972	4	30.524	447	20.508	.000
¿Aproximadamente cuántas horas busca información académica en internet cada mes?	22398.406	4	33.108	447	676.524	.000

Las pruebas F sólo se deben utilizar con fines descriptivos porque los clústeres se han elegido para maximizar las diferencias entre los casos de distintos clústeres. Los niveles de significación observados no están corregidos para esto y, por lo tanto, no se pueden interpretar como pruebas de la hipótesis de que los medias de clúster son iguales.

Número de casos en cada clúster		
Clúster	1	147.000
	2	79.000
	3	59.000
	4	53.000
	5	114.000
Válido		452.000
Perdidos		.000

Anexo 12: Uso de internet en actividades de entretenimiento clúster de dos grupos

Centros de clústeres finales		
	Clúster	
	1	2
¿Aproximadamente cuántas horas a la semana chatea por diversión?	6.74	19.21
¿Aproximadamente cuántas horas a la semana utiliza redes sociales?	7.96	23.60
¿Aproximadamente cuántas horas a la semana utiliza juegos en línea?	1.78	6.31
¿Aproximadamente cuántas horas a la semana descarga música, videos y programas?	3.96	6.69
¿Aproximadamente cuántos videos para entretenimiento mira en youtube cada semana?	6.06	14.31

ANOVA						
	Clúster		Error		F	Sig.
	Media cuadrática	gl	Media cuadrática	gl		
¿ Aproximadamente cuántas horas a la semana chatea por diversión?	13782.386	1	27.021	450	510.056	.000
¿Aproximadamente cuántas horas a la semana utiliza redes sociales?	21658.716	1	39.435	450	549.229	.000
¿Aproximadamente cuántas horas a la semana utiliza juegos en línea?	1819.587	1	26.432	450	68.840	.000
¿Aproximadamente cuántas horas a la semana descarga música, videos y programas?	656.623	1	14.959	450	43.894	.000
¿Aproximadamente cuántos videos para entretenimiento mira en youtube cada semana?	6036.213	1	45.713	450	132.046	.000

Las pruebas F sólo se deben utilizar con fines descriptivos porque los clústeres se han elegido para maximizar las diferencias entre los casos de distintos clústeres. Los niveles de significación observados no están corregidos para esto y, por lo tanto, no se pueden interpretar como pruebas de la hipótesis de que las medias de clúster son iguales.

Número de casos en cada clúster		
Clúster	1	331.000
	2	121.000
Válido		452.000
Perdidos		.000

Anexo 13: Uso de internet en actividades de entretenimiento clúster de tres grupos

Centros de clústeres finales			
	Clúster		
	1	2	3
¿Aproximadamente cuántas horas a la semana chatea por diversión?	20.47	12.06	6.21
¿Aproximadamente cuántas horas a la semana utiliza redes sociales?	27.92	11.37	7.61
¿Aproximadamente cuántas horas a la semana utiliza juegos en línea?	4.71	7.13	1.00
¿Aproximadamente cuántas horas a la semana descarga música, videos y programas?	6.52	6.84	3.37
¿Aproximadamente cuántos videos para entretenimiento mira en youtube cada semana?	11.86	16.75	4.15

ANOVA						
	Clúster		Error		F	Sig.
	Media cuadrática	gl	Media cuadrática	gl		
¿Aproximadamente cuántas horas a la semana chatea por diversión?	6711.599	2	27.881	449	240.719	.000
¿Aproximadamente cuántas horas a la semana utiliza redes sociales?	13147.311	2	29.198	449	450.287	.000
¿Aproximadamente cuántas horas a la semana utiliza juegos en línea?	1496.338	2	23.878	449	62.665	.000
¿Aproximadamente cuántas horas a la semana descarga música, videos y programas?	598.337	2	13.790	449	43.390	.000
¿Aproximadamente cuántos videos para entretenimiento mira en YouTube cada semana?	6330.962	2	31.058	449	203.842	.000

Las pruebas F sólo se deben utilizar con fines descriptivos porque los clústeres se han elegido para maximizar las diferencias entre los casos de distintos clústeres. Los niveles de significación observados no están corregidos para esto y, por lo tanto, no se pueden interpretar como pruebas de la hipótesis de que las medias de clúster son iguales.

Número de casos en cada clúster		
Clúster	1	83.000
	2	97.000
	3	272.000
Válido		452.000
Perdidos		.000

Anexo 14: Uso de internet en actividades de entretenimiento clúster de cuatro grupos

Centros de clústeres finales				
	Clúster			
	1	2	3	4
¿Aproximadamente cuántas horas a la semana chatea por diversión?	15.32	19.88	6.24	10.79
¿Aproximadamente cuántas horas a la semana utiliza redes sociales?	13.74	28.04	7.47	10.60
¿Aproximadamente cuántas horas a la semana utiliza juegos en línea?	17.41	3.59	1.02	2.85
¿Aproximadamente cuántas horas a la semana descarga música, videos y programas?	6.15	6.40	3.35	7.18
¿Aproximadamente cuántos videos para entretenimiento mira en youtube cada semana?	11.97	11.49	4.10	18.97

ANOVA						
	Clúster		Error		F	Sig.
	Media cuadrática	gl	Media cuadrática	gl		
¿Aproximadamente cuántas horas a la semana chatea por diversión?	4265.305	3	29.344	448	145.356	.000
¿Aproximadamente cuántas horas a la semana utiliza redes sociales?	8934.586	3	28.126	448	317.659	.000
¿Aproximadamente cuántas horas a la semana utiliza juegos en línea?	2713.153	3	12.443	448	218.044	.000
¿Aproximadamente cuántas horas a la semana descarga música, videos y programas?	403.688	3	13.788	448	29.277	.000
¿Aproximadamente cuántos videos para entretenimiento mira en YouTube cada semana?	4586.464	3	28.678	448	159.930	.000

Las pruebas F sólo se deben utilizar con fines descriptivos porque los clústeres se han elegido para maximizar las diferencias entre los casos de distintos clústeres. Los niveles de significación observados no están corregidos para esto y, por lo tanto, no se pueden interpretar como pruebas de la hipótesis de que las medias de clúster son iguales.

Número de casos en cada clúster		
Clúster	1	34.000
	2	82.000
	3	268.000
	4	68.000
Válido		452.000
Perdidos		.000

Anexo 15: Uso de internet en actividades de entretenimiento clúster de cinco grupos

Centros de clústeres finales					
	Clúster				
	1	2	3	4	5
¿Aproximadamente cuántas horas a la semana chatea por diversión?	14.77	6.18	20.00	9.19	19.27
¿Aproximadamente cuántas horas a la semana utiliza redes sociales?	13.03	7.35	27.76	9.24	25.94
¿Aproximadamente cuántas horas a la semana utiliza juegos en línea?	17.06	1.05	6.44	2.30	2.90
¿Aproximadamente cuántas horas a la semana descarga música, videos y programas?	6.23	3.32	6.88	7.09	5.83
¿Aproximadamente cuántos videos para entretenimiento mira en YouTube cada semana?	10.23	3.87	24.18	17.45	6.92

ANOVA						
	Clúster		Error		F	Sig.
	Media cuadrática	gl	Media cuadrática	gl		
¿Aproximadamente cuántas horas a la semana chatea por diversión?	3326.314	4	28.270	447	117.662	.000
¿Aproximadamente cuántas horas a la semana utiliza redes sociales?	6694.217	4	28.249	447	236.968	.000
¿Aproximadamente cuántas horas a la semana utiliza juegos en línea?	1887.210	4	13.792	447	136.831	.000
¿Aproximadamente cuántas horas a la semana descarga música, videos y programas?	297.323	4	13.868	447	21.440	.000
¿Aproximadamente cuántos videos para entretenimiento mira en youtube cada semana?	4865.009	4	15.989	447	304.274	.000

Las pruebas F sólo se deben utilizar con fines descriptivos porque los clústeres se han elegido para maximizar las diferencias entre los casos de distintos clústeres. Los niveles de significación observados no están corregidos para esto y, por lo tanto, no se pueden interpretar como pruebas de la hipótesis de que los medias de clúster son iguales.

Número de casos en cada clúster		
Clúster	1	31.000
	2	257.000
	3	34.000
	4	67.000
	5	63.000
Válido		452.000
Perdidos		.000

Anexo 16: Análisis discriminante para la verificación del porcentaje de exactitud en actividades académicas

Resultados de clasificación ^a					
		Número de caso de clúster	Pertenencia a grupos pronosticada		Total
			1	2	
Original	Recuento	1	322	1	323
		2	2	127	129
	%	1	99.7	.3	100.0
		2	1.6	98.4	100.0

a. 99,3% de casos agrupados originales clasificados correctamente.

Resultados de clasificación ^a						
		Número de caso de clúster	Pertenencia a grupos pronosticada			Total
			1	2	3	
Original	Recuento	1	207	0	8	215
		2	0	73	0	73
		3	3	0	161	164
	%	1	96.3	.0	3.7	100.0
		2	.0	100.0	.0	100.0
		3	1.8	.0	98.2	100.0

a. 97,6% de casos agrupados originales clasificados correctamente.

Resultados de clasificación ^a							
		Número de caso de clúster	Pertenencia a grupos pronosticada				Total
			1	2	3	4	
Original	Recuento	1	87	0	0	6	93
		2	0	73	0	0	73
		3	2	0	168	1	171
		4	3	0	3	109	115
	%	1	93.5	.0	.0	6.5	100.0
		2	.0	100.0	.0	.0	100.0
		3	1.2	.0	98.2	.6	100.0
		4	2.6	.0	2.6	94.8	100.0

a. 96,7% de casos agrupados originales clasificados correctamente.

Resultados de clasificación ^a								
		Número de caso de clúster	Pertenencia a grupos pronosticada					Total
			1	2	3	4	5	
Original	Recuento	1	144	2	0	0	1	147
		2	1	77	0	0	1	79
		3	0	0	58	1	0	59
		4	0	0	1	52	0	53
		5	1	0	0	2	111	114
	%	1	98.0	1.4	.0	.0	.7	100.0
		2	1.3	97.5	.0	.0	1.3	100.0
		3	.0	.0	98.3	1.7	.0	100.0
		4	.0	.0	1.9	98.1	.0	100.0
		5	.9	.0	.0	1.8	97.4	100.0

a. 97,8% de casos agrupados originales clasificados correctamente.

Anexo 17: Análisis discriminante para la verificación del porcentaje de exactitud en actividades de entretenimiento

Resultados de clasificación ^a					
		Número de caso de clúster	Pertenencia a grupos pronosticada		Total
			1	2	
Original	Recuento	1	328	3	331
		2	8	113	121
	%	1	99.1	.9	100.0
		2	6.6	93.4	100.0

a. 97,6% de casos agrupados originales clasificados correctamente.

Resultados de clasificación ^a						
		Número de caso de clúster	Pertenencia a grupos pronosticada			Total
			1	2	3	
Original	Recuento	1	80	3	0	83
		2	1	93	3	97
		3	3	1	268	272
	%	1	96.4	3.6	.0	100.0
		2	1.0	95.9	3.1	100.0
		3	1.1	.4	98.5	100.0

a. 97,6% de casos agrupados originales clasificados correctamente.

Resultados de clasificación ^a							
		Número de caso de clúster	Pertenencia a grupos pronosticada				Total
			1	2	3	4	
Original	Recuento	1	31	1	0	2	34
		2	4	78	0	0	82
		3	2	3	258	5	268
		4	0	0	4	64	68
	%	1	91.2	2.9	.0	5.9	100.0
		2	4.9	95.1	.0	.0	100.0
		3	.7	1.1	96.3	1.9	100.0
		4	.0	.0	5.9	94.1	100.0

a. 95,4% de casos agrupados originales clasificados correctamente.

Resultados de clasificación ^a								
		Número de caso de clúster	Pertenencia a grupos pronosticada					Total
			1	2	3	4	5	
Original	Recuento	1	29	0	1	0	1	31
		2	2	253	0	1	1	257
		3	0	0	32	1	1	34
		4	0	2	2	63	0	67
		5	1	1	0	0	61	63
	%	1	93.5	.0	3.2	.0	3.2	100.0
		2	.8	98.4	.0	.4	.4	100.0
		3	.0	.0	94.1	2.9	2.9	100.0
		4	.0	3.0	3.0	94.0	.0	100.0
		5	1.6	1.6	.0	.0	96.8	100.0

a. 96,9% de casos agrupados originales clasificados correctamente.

Anexo 18: Modelo de Regresión Logística Binaria aplicada al perfil de uso de internet en actividades académicas

Codificación de variable dependiente	
Valor original	Valor interno
1	0
2	1

Codificaciones de variables categóricas						
		Frecuencia	Codificación de parámetro			
			(1)	(2)	(3)	(4)
Los ingresos mensuales de su familia son de:	Hasta 350 dólares	157	1,000	,000	,000	,000
	Hasta 600 dólares	150	,000	1,000	,000	,000
	Hasta 1000 dólares	87	,000	,000	1,000	,000
	Hasta 1.500 dólares	36	,000	,000	,000	1,000
	Más de 1.500 dólares	22	,000	,000	,000	,000

Pruebas ómnibus de coeficientes de modelo				
		Chi-cuadrado	gl	Sig.
Paso 1	Escalón	1,322	4	,858
	Bloque	1,322	4	,858
	Modelo	1,322	4	,858

Resumen del modelo			
Escalón	Logaritmo de la verosimilitud -2	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	539,252 ^a	,003	,004

a. La estimación ha terminado en el número de iteración 4 porque las estimaciones de parámetro han cambiado en menos de ,001.

Prueba de Hosmer y Lemeshow			
Escalón	Chi-cuadrado	gl	Sig.
1	,000	3	1,000

Tabla de contingencia para la prueba de Hosmer y Lemeshow						
		Número de caso de clúster = 1		Número de caso de clúster = 2		Total
		Observado	Esperado	Observado	Esperado	
Paso 1	1	28	28,000	8	8,000	36
	2	64	64,000	23	23,000	87
	3	16	16,000	6	6,000	22
	4	111	111,000	46	46,000	157
	5	104	104,000	46	46,000	150

Anexo 19: Modelo de Regresión Logística Multinomial aplicada al perfil de uso de internet en actividades de entretenimiento

Información de ajuste de los modelos						
Modelo	Criterios de ajuste de modelo			Contraste de la razón de verosimilitud		
	AIC	BIC	Logaritmo de la verosimilitud -2	Chi-cuadrado	gl	Sig.
Sólo interceptación	66,658	74,885	62,658			
Final	62,598	103,735	42,598	20,060	8	,010

Bondad de ajuste			
	Chi-cuadrado	gl	Sig.
Pearson	,000	0	.
Desvianza	,000	0	.

Pseudo R cuadrado	
Cox y Snell	,043
Nagelkerke	,051
McFadden	,023

Contraste de la razón de verosimilitud						
Efecto	Criterios de ajuste de modelo			Contraste de la razón de verosimilitud		
	AIC de modelo reducido	BIC de modelo reducido	Logaritmo de la verosimilitud -2 de modelo reducido	Chi-cuadrado	gl	Sig.
Interceptación	62,598	103,735	42,598 ^a	,000	0	.
ing	66,658	74,885	62,658	20,060	8	,010

El estadístico de chi-cuadrado es la diferencia de los logaritmos de la verosimilitud -2 entre el modelo final y el modelo reducido. El modelo reducido se forma omitiendo un efecto del modelo final. La hipótesis nula es que todos los parámetros de dicho efecto son 0.

a. Este modelo reducido es equivalente al modelo final porque omitir el efecto no aumenta los grados de libertad.

Anexo 20: Modelo de Regresión Logística Binaria aplicada al uso de internet en actividades académicas sobre el rendimiento académico

Codificación de variable dependiente	
Valor original	Valor interno
,00	0
1,00	1

Pruebas ómnibus de coeficientes de modelo				
		Chi-cuadrado	gl	Sig.
Paso 1	Escalón	1,453	1	,228
	Bloque	1,453	1	,228
	Modelo	1,453	1	,228

Prueba de Hosmer y Lemeshow			
Escalón	Chi-cuadrado	gl	Sig.
1	,000	0	.

Tabla de clasificación ^a					
Observado			Pronosticado		
			rendimiento_academico		Corrección de porcentaje
			,00	1,00	
Paso 1	rendimiento_academico	,00	350	0	100,0
		1,00	102	0	,0
Porcentaje global					77,4

a. El valor de corte es ,500

Anexo 21: Modelo de Regresión Logística Binaria aplicada al uso de internet en actividades de entretenimiento sobre el rendimiento académico

Codificación de variable dependiente	
Valor original	Valor interno
,00	0
1,00	1

Pruebas ómnibus de coeficientes de modelo				
		Chi-cuadrado	gl	Sig.
Paso 1	Escalón	,274	1	,601
	Bloque	,274	1	,601
	Modelo	,274	1	,601

Prueba de Hosmer y Lemeshow			
Escalón	Chi-cuadrado	gl	Sig.
1	,926	1	,336

Tabla de clasificación ^a					
		Pronosticado			
		rendimiento_academico		Corrección de porcentaje	
Observado		,00	1,00		
Paso 1	rendimiento_academico	,00	350	0	100,0
		1,00	102	0	,0
Porcentaje global					77,4

a. El valor de corte es ,500

Anexo 22: Objetivos proporcionados por el director del proyecto

- Determinar los usos de internet en las universidades de categorías A, B y C
- Relacionar los usos de internet con el rendimiento académico y con los niveles de ingreso.
- Establecer una línea de base respecto al tema en el país.
- Publicar el texto “La educación virtual en Ecuador”.