



UNIVERSIDAD TÉCNICA PARTICULAR DE LOJA

La Universidad Católica de Loja

ÁREA TÉCNICA

TÍTULO DE INGENIERO EN INFORMÁTICA

La educación virtual en el Ecuador:

**Análisis de los usos de la tecnología en los estudiantes de la Escuela
Superior Politécnica del Litoral**

TRABAJO DE TITULACIÓN

AUTOR: Caicedo Vera, Isaac Augusto

DIRECTOR: Torres Díaz, Juan Carlos, Phd

CODIRECTOR: Correa Granda, Carlos Aníbal, Mgs

CENTRO UNIVERSITARIO GUAYAQUIL

2016

APROBACIÓN DEL DIRECTOR DE TRABAJO DE TITULACIÓN

Dr.

Ing. Torres Díaz, Juan Carlos

DOCENTE DE LA TITULACIÓN

De mi consideración:

El presente trabajo de titulación: Análisis de los usos de tecnología en los estudiantes de la Escuela Superior Politécnica del Litoral, realizado por Caicedo Vera Isaac Augusto, ha sido orientado y revisado durante su ejecución, por cuanto se aprueba la presentación del mismo.

Loja, 11 de febrero del 2016

f)

DECLARACIÓN DE AUTORÍA Y CESIÓN DE DERECHOS

“Yo Caicedo Vera Isaac Augusto declaro ser autor del presente trabajo de titulación: Análisis de los usos de la Tecnología en los Estudiantes de la Escuela Superior Politécnica del Litoral, de la titulación de Ingeniería Informática, siendo el Dr. Juan Carlos Torres Díaz, director del presente trabajo; y eximo expresamente a la Universidad Técnica Particular de Loja y a sus representantes legales de posibles reclamos o acciones legales. Además certifico que las ideas, conceptos, procedimientos y resultados vertidos en el presente trabajo investigativo, son de mi exclusiva responsabilidad.

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

f)

Autor Caicedo Vera Isaac Augusto

Cédula 1307048908

DEDICATORIA

A la Señora del Rosario, quien desde mi expresión de fe, es la Madre que he consagrado los momentos de mi vida, sueños, deseos y proyectos.

A la Hna. Bertha de Lourdes Serrano Rodas, O.P., mujer de luz; por su apoyo e incondicional motivación para terminar este proyecto de tesis y constituirse en el pilar de mi vida académica, profesional y laboral.

AGRADECIMIENTO

Gratitud imperecedera a mis padres y maestros de vida: Jaime Eduardo y Ana María, por su apoyo incondicional en mis momentos de luces y sombras y desde la imperfección humana han cumplido perfectamente su rol de padres y compañeros de camino.

A mis abuelos Carmen e Isaac, mis ángeles, que en paz descansen. En vida, no dejaron de confiar en mí y desde el cielo son mis ángeles guardianes. Su esfuerzo y arduo trabajo, permitió que pudiera empezar mis estudios en la Universidad.

A mi abuela Nelly Rivero, mujer sin par, aquella guerrera que con sentimientos afectivos y efectivos es parte de este triunfo.

A mis hermanas, María y Nelly Caicedo Vera en especial, por darme a 3 seres maravillosos que son la fuerza para continuar y mejorar.

A Paola Noriega, por ser quien ha inducido momentos de sueños y metas, y su tesón es ejemplo para crecer profesionalmente.

A Ma. Isabel Sandoval, gran ayuda en mis momentos de necesidad.

A Stefanie Hormaza y Marjorie Jiménez por su ayuda en la aplicación de encuestas del presente proyecto.

Al Phd. Juan Carlos Torres Díaz, gran maestro que ha sabido mostrarnos el camino y la guía en este Proyecto.

A la Lcda. Lidia Villacís, quien motivó mi trabajo y animó desde Loja el cumplimiento de mis metas en la UTPL.

A mis compañeros y colegas Tatiana Jaramillo, Jéssica Jima y Carlos Luis Manrique.

A los docentes y personal de la UTPL, quienes desde la cátedra y la oficina, han logrado que en este trabajo queden plasmadas las enseñanzas impartidas en la Modalidad a Distancia.

ÍNDICE DE CONTENIDOS

APROBACIÓN DEL DIRECTOR DE TRABAJO DE FIN DE TITULACIÓN	ii
DECLARACIÓN DE AUTORÍA Y CESIÓN DE DERECHOS	iii
DEDICATORIA	iv
AGRADECIMIENTO	v
ÍNDICE DE CONTENIDOS	vi
ÍNDICE DE TABLAS.....	ix
ÍNDICE DE FIGURAS	x
RESUMEN EJECUTIVO	1
ABSTRACT.....	2
CAPÍTULO I: INTRODUCCIÓN	3
1.1 Introducción.....	4
1.2 Objetivos.	5
1.3 Preguntas de Investigación.	5
1.4 Hipótesis.	5
1.5 Estructura del documento.....	6
CAPÍTULO II: MARCO TEÓRICO	7
2.1 Antecedentes.	8
2.1.1 La brecha digital.	8
2.1.2 Factores que influyen en la medición de la brecha digital.....	10
2.1.2.1 <i>Internet y el logro académico.</i>	10
2.1.2.2 <i>Internet y las actividades de entretenimiento.</i>	11
2.1.2.3 <i>El nivel de ingresos y el uso de Internet.</i>	12
2.1.3 Resultados de medición de la brecha digital a nivel mundial.	12
2.2 Acceso a las Tecnologías de la Información y Comunicaciones en el Ecuador.	16
2.3 Minería de datos.	17
2.3.1 Operaciones y Técnicas de minería de datos.	17
2.3.2 Aplicaciones de la minería de datos.	19
2.3.3 Minería de Datos en la Educación (MDE), Educational Data Mining (EDM).	20
2.3.4 Procedimiento para descubrir la información oculta en bases de datos o KDD.	24
2.3.4.1 <i>Limpieza e integración.</i>	25
2.3.4.2 <i>Selección y transformación.</i>	25

2.3.4.3	<i>Minería de Datos</i>	25
2.3.4.4	<i>Evaluación y presentación</i>	26
2.3.4.5	<i>Presentación del Conocimiento</i>	26
CAPÍTULO III: METODOLOGÍA DE INVESTIGACIÓN.....		27
3.1	Población y Tamaño de la muestra.....	28
3.2	Procedimiento para descubrir el conocimiento en el Proyecto de Investigación (KDD).....	29
3.2.1	Recolección de Información.....	29
3.2.1.1	<i>Descripción del contenido del instrumento de recolección de información</i> ...29	
3.2.2	Procedimiento para la fase de Selección y Transformación de los datos.	31
3.2.3	Minería de Datos.	32
3.2.3.1	<i>Análisis de Correspondencia Simple ACS</i>	32
3.2.3.2	<i>χ² - chi cuadrado</i>	33
3.2.3.3	<i>Análisis Factorial</i>	33
3.2.3.4	<i>Análisis de conglomerados o análisis clúster</i>	35
3.2.3.5	<i>Análisis Discriminante</i>	36
3.2.3.6	<i>Análisis de Varianza</i>	37
3.2.3.7	<i>Regresión Logística</i>	37
3.2.4	Procedimiento para Evaluación y presentación de datos.	38
CAPÍTULO IV: MUESTRA DE RESULTADOS		39
4.1	Generalidades sobre la muestra.	40
4.1.1	Datos relevantes sobre la población: Género, edades, ingresos familiares.	42
4.1.2	Datos relevantes sobre el conocimiento y el uso de Internet.....	44
4.2	Relaciones entre variables.	47
4.3	Perfil uso de Internet para actividades académicas.	48
4.4	Perfil uso de Internet para actividades de entretenimiento.....	52
4.5	Fase de Evaluación.	54
4.5.1	Evaluación del perfil uso de Internet y actividades académicas.....	55
4.5.2	Evaluación del perfil de uso de Internet para actividades de entretenimiento..	56
4.5.3	Evaluación de la incidencia del uso de Internet para actividades académicas sobre el rendimiento académico.....	58
4.5.4	Evaluación de la incidencia del uso de Internet para actividades de entretenimiento sobre el rendimiento académico.....	59
CAPÍTULO V: DISCUSIÓN DE RESULTADOS		61
5.1	Discusión de resultados.....	62
CAPÍTULO VI: CONCLUSIONES Y RECOMENDACIONES.....		65
6.1	Conclusiones.....	66

6.2	Recomendaciones y trabajos futuros.....	67
	BIBLIOGRAFÍA.....	69
	ANEXOS.....	73
	Anexo 1: Instrumento de recolección de información.....	74
	Anexo 2: Análisis factorial de las variables relacionadas con actividades académicas de los estudiantes.....	76
	Anexo 3: Análisis factorial de las variables relacionadas con actividades de entretenimiento de los estudiantes.....	77
	Anexo 4: Análisis del Clúster en dos grupos según el uso de Internet en actividades académicas de los estudiantes.....	78
	Anexo 5: Análisis del Clúster en tres grupos según el uso de Internet en actividades académicas de los estudiantes.....	80
	Anexo 6: Análisis del Clúster en cuatro grupos según el uso de Internet en actividades académicas de los estudiantes.....	82
	Anexo 7: Análisis del Clúster en cinco grupos según el uso de Internet en actividades académicas de los estudiantes.....	84
	Anexo 8: Análisis Discriminante para probar el porcentaje de exactitud de actividades académicas.....	86
	Anexo 9: Análisis del Clúster en dos grupos según el uso de Internet en actividades de entretenimiento.....	88
	Anexo 10: Análisis del Clúster en tres grupos según el uso de Internet en actividades de entretenimiento.....	90
	Anexo 11: Análisis del Clúster en cuatro grupos según el uso de Internet en actividades de entretenimiento.....	92
	Anexo 12: Análisis del Clúster en cinco grupos según el uso de Internet en actividades de entretenimiento.....	94
	Anexo 13: Análisis Discriminante para probar el porcentaje de exactitud de actividades de entretenimiento.....	96
	Anexo 14: Regresión Logística Binaria aplicada al perfil de Uso de Internet y actividades académicas.....	98
	Anexo 15: Regresión Logística Multinomial aplicada al perfil de Uso de Internet y actividades de entretenimiento.....	100
	Anexo 16: Regresión Logística Binaria aplicada al Uso de Internet y actividades académicas sobre el rendimiento académico.....	101
	Anexo 17: Regresión Logística Binaria aplicada al Uso de Internet y actividades de entretenimiento sobre el rendimiento académico.....	103

ÍNDICE DE TABLAS

Tabla 1: “Operaciones de minería de datos y sus técnicas asociadas”	18
Tabla 2: Diferencias entre Minería de Datos y Minería de Datos en la Educación	22
Tabla 3: Usuarios de MDE / Beneficiarios	22
Tabla 4: Oferta Académica de la ESPOL	40
Tabla 5: Horas de conexión de los estudiantes.....	45
Tabla 6: Años de experiencia utilizando Internet.....	46
Tabla 7: Accesos semanales a la plataforma virtual de la Universidad.....	47
Tabla 8: Uso de Internet en Actividades Académicas	49
Tabla 9: Clasificación del Discriminante grupos de Actividades Académicas	50
Tabla 10: Uso de Internet en Actividades de Entretenimiento	52
Tabla 11: Clasificación del Discriminante grupos de Actividades de Entrenamiento	53
Tabla 12: Coeficientes del modelo de regresión logística para el perfil de uso de Internet y actividades académicas	56
Tabla 13: Coeficientes del modelo de regresión logística multinomial para el perfil de Uso de Internet para actividades de entretenimiento.....	57
Tabla 14: Coeficientes de la regresión logística binaria entre el Uso de Internet en actividades académicas sobre el rendimiento académico	59
Tabla 15: Coeficientes de la regresión logística binaria entre el Uso de Internet en actividades de entretenimiento sobre el rendimiento académico	60

ÍNDICE DE FIGURAS

Figura 1: Índice de Desarrollo de las TIC 2012 y 2013	14
Figura 2: Networked Readiness Index Ecuador 2014	15
Figura 3: Evolución del Concepto de Minería de Datos	17
Figura 4: Técnicas de Minería de Datos	18
Figura 5: Metodología para el descubrimiento de conocimiento en bases de datos.....	24
Figura 6: Proceso de obtención del conocimiento KDD	26
Figura 7: Porcentaje de estudiantes por género.....	42
Figura 8: Distribución de edades	43
Figura 9: Ingresos mensuales de los familiares de los encuestados.....	43
Figura 10: Lugar de Conexión.....	44
Ilustración 11: Días a la semana que se conecta a Internet	44
Figura 12: Nivel de conocimiento en el uso de Internet	45
Figura 13: Uso de Internet en Actividades Académicas.....	49
Figura 14: Perfil de uso de Internet para Actividades Académicas.....	51
Figura 15: Uso de Internet en Actividades de Entretenimiento	52
Figura 16: Perfil de uso de Internet para Actividades de Entretenimiento.....	54

RESUMEN EJECUTIVO

Este trabajo, cuya línea de investigación es la Tecnología aplicada a la educación, permite analizar las relaciones entre los niveles de ingreso de los familiares de los estudiantes de la Escuela Superior Politécnica del Litoral (ESPOL) y el uso de Internet en actividades académicas y de entretenimiento. De igual manera, mide las relaciones entre el rendimiento académico y el uso de Internet en actividades académicas y de entretenimiento.

El desarrollo de la presente investigación se realizó aplicando el proceso de KDD, de la Minería de Datos y el uso del software SPSS versión 19.

Se identifica que no se cumplen las hipótesis propuestas, puesto que no existe relación entre el nivel de ingresos económicos de los familiares de los estudiantes y el uso de Internet para el aprendizaje; tampoco existe relación entre el nivel de ingresos y el uso de Internet para el entretenimiento. El uso de la tecnología en el aprendizaje no incide en el rendimiento académico, así como el uso de la tecnología en para el entretenimiento no incide en el rendimiento académico.

PALABRAS CLAVE:

Brecha digital, Minería de Datos, Minería de Datos en la Educación, Tecnologías de la Información y Comunicación, KDD, Internet.

ABSTRACT

This work, whose research line is Technology Applied to Education, analyze the relationship between income levels of the families of the students of the Escuela Superior Politecnica del Litoral (ESPOL) and the use of Internet in academic activities and entertainment. In the same way, analyze the measures of the relationship between academic performance and the Internet use in academic activities and entertainment.

The development of this research was done by applying the KDD process using Data Mining and the use of SPSS version 19 software.

Finally, identifies the proposed hypotheses are not fulfilled because there is no relationship between the level of income of the families of the students and the use of the Internet for learning, and there is no relationship between the level of income and the use of Internet for entertainment. The use of technology in learning does not affect academic performance, and the use of technology for entertainment does not affect academic performance.

KEYWORDS:

Digital Divide, Data Mining, Educational Data Mining, Information and Communications Technology, KDD, Internet.

CAPÍTULO I: INTRODUCCIÓN

1.1 Introducción.

El Siglo XXI se ha caracterizado por la rápida evolución y revolución de las tecnologías de la información y la comunicación. Es el siglo de la alfabetización digital, que debe responder a los nuevos desafíos planteados por la sociedad de la información y del conocimiento (Avello Martínez, y otros, 2013), no solo se trata de acceder a la información o utilizar las tecnologías y dispositivos, sino con la identificación de las competencias tecnológicas e informáticas necesarias para acceder a la información.

Nuestro país ha iniciado una lucha para disminuir la brecha digital, según (Caicedo, 2014) Ecuador en los últimos 5 años ha ascendido 32 puestos en los rankings del Global Information Technology Report (GITR), nuestra sociedad se ha alfabetizado digitalmente. El gobierno del Presidente Ec. Rafael Correa ha manifestado su interés en disminuir los niveles de analfabetismo, ya sea en educación, como en las competencias de la alfabetización digital, disminuyendo y avanzando en la reducción de la brecha digital, a través de los retos de: duplicar la penetración y velocidad de Banda Ancha Fija, duplicar la penetración de Banda Ancha Móvil, reestructurar ARCOTEL, mejorar la calidad del servicio móvil, duplicar los ciudadanos alfabetizados digitalmente (Erazo, 2015).

Considerarse alfabeto digital, lleva consigo 6 nuevas habilidades relacionadas con las TICs (Cobo, 2011):

- **Leer en pantalla:** Contexto en el que se publican nuevos contenidos.
- **Interactuar:** Interactuamos de manera virtual.
- **Compartir:** Como seres sociales nuestra actividad se entrecruza en bibliotecas compartidas.
- **Acceder:** lograr acceso a la información en cualquier momento.
- **Fluir:** Los datos circulan exponencialmente. Twitter, Facebook, medios RSS, blogs, geolocalización, etc.
- **Generar:** Generación de contenidos en nuevos formatos y a través de nuevos medios.

El presente trabajo de investigación, forma parte del Proyecto: **“La educación virtual en el Ecuador. Usos de la tecnología en la universidad ecuatoriana”**. Y se convertirá en un aporte y referente para analizar y determinar los principales usos de la tecnología en las universidades ecuatorianas.

Esta tesis, tiene como propósito identificar y analizar los usos de Internet en la Escuela Superior Politécnica del Litoral (ESPOL) de la ciudad de Guayaquil, así como la definición de los perfiles de uso de Internet en actividades académicas y de entretenimiento de los estudiantes de la ESPOL.

1.2 Objetivos.

El presente proyecto tiene como objetivos:

- 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”¹.

1.3 Preguntas de Investigación.

Pregunta 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?

Pregunta 2: ¿Cómo se relaciona el rendimiento académico y los usos de Internet en actividades académicas y de aprendizaje?

1.4 Hipótesis.

¹ Proyecto: La educación virtual en el Ecuador. Uso de la tecnología en la universidad ecuatoriana.

Relacionadas con la 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 con la 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 Estructura del documento.

El documento está estructurado en 5 capítulos, el **Capítulo 1: Introducción** presenta la introducción, objetivos del proyecto, preguntas de investigación e hipótesis.

El **Capítulo 2: Marco Teórico**, describe la información sobre los conceptos de brecha digital, factores que influyen en la brecha digital, minería de datos, proceso KDD.

El **Capítulo 3: Metodología de investigación**, que incluye los aspectos metodológicos realizados para obtener la información de estudio del presente trabajo.

Capítulo 4: Análisis y Muestra de resultados, que corresponde al análisis de los resultados obtenidos del proceso metodológico de investigación.

Capítulo 5: Discusión de Resultados, que corresponde a la discusión de los resultados obtenidos y el contraste con trabajos relacionados.

Finalmente, el **Capítulo 6: Conclusiones y Recomendaciones** que presenta el cierre del presente trabajo y las conclusiones frente a las hipótesis, además se incluye sugerencias sobre la presente investigación y trabajos futuros.

CAPÍTULO II: MARCO TEÓRICO

2.1 Antecedentes.

La historia del ser humano, está marcada por la evolución y la revolución, esto significa que, desde su razón antropológica el ser humano es un ser social, por ende curioso Y comunicativo, descubridor de cosas nuevas, evolutivo en el tiempo y que cambia su comportamiento de acuerdo a la cultura y situación en la que se encuentre. El descubrimiento de las cosas desde los inicios de la civilización, por ejemplo en la prehistoria: el fuego, la piedra, herramientas rústicas, armas y artefactos prehistóricos, es reflejo de esta innata curiosidad a la que estamos inmersos en nuestro proceso evolutivo.

En la actualidad, la globalización nos invita a estar conectados, vivimos en un mundo competitivo que nos mueve a estar comunicados en su totalidad, sin embargo; no todos tenemos acceso a los mismos recursos tecnológicos y técnicos, por ejemplo: teléfono, teléfono celular, computador, teléfono celular inteligente, Internet, todavía existen limitaciones de acceso a estos medios y recursos.

Desde su aparición en los años 60, Internet se ha mostrado como el gigante que comunica y divide al mundo, esta red de redes fuertemente ha ingresado al mercado y a la mayoría de los hogares del mundo. Al ingresar en el mundo ha generado una brecha, es decir una pequeña separación entre los que saben y los que no saben utilizarla, los que tienen el acceso y los que no, los que tienen los recursos y los que no, entre los que la utilizan correctamente y los que apenas saben utilizarla, los que tienen ingresos económicos y mejores posibilidades y los que no. El término que mejor acuña esta desigualdad es el de **“brecha digital”**.

2.1.1 La brecha digital.

Clarifiquemos algunos conceptos y autores, puesto que es importante definir desde qué ángulo se quiere analizar la brecha digital, debido a que en la actualidad se expresa en torno a: brecha de género, cultura, de poder, de nivel económico, de acceso, educación, tecnología, etc., en nuestra sociedad de la información y del conocimiento:

Algunas definiciones y conceptos sobre brecha digital:

“Brecha digital es el vacío que existe entre los individuos y las comunidades que tienen acceso a las tecnologías de la información y aquellos que no la tienen”.

(Molinari, 2011)

“La brecha digital puede ser definida en términos de desigualdad de posibilidades que existen para acceder a la información, al conocimiento y la educación mediante las nuevas tecnologías de información” (Serrano Santoyo & Martínez Martínez, 2003).

De aquí es importante recalcar 3 factores que intervienen en la utilización y conceptualización de la brecha digital:

- **Infraestructura:** equipos conectados en red.
- **Acceso:** acceso a la información y conocimiento.
- **Habilidades y competencias digitales:** uso de las herramientas y tecnologías de información (Martínez, Ascencio, & Serrano, La Brecha Digital, 2005).

Además, según (Molinari, 2011), la Brecha Digital sucede por 3 condiciones básicas:

1. Las personas no tienen acceso a la tecnología, porque no pueden cubrir con sus gastos.
2. No saben utilizar los recursos y medios de las TIC.
3. No conocen los beneficios de la tecnología.

Por eso se infiere que “La brecha digital se refiere a la ruptura que existe entre individuos y sociedades que tienen recursos para participar en la era de la información y aquellos que carecen de ellos”.

Para (Castaño M., 2010) existe un acuerdo para definir las dimensiones de la brecha digital:

- El acceso a Internet: motivación, infraestructura y lugar de conexión.
- Alfabetización digital: diferencias en las habilidades de uso de Internet.

- Tiempo en la red: Intensidad de uso de Internet.
- Adopción de la Tecnología: finalidades de uso de Internet.

2.1.2 Factores que influyen en la medición de la brecha digital.

Según (Martínez & Serrano, La Brecha Digital, 2007) para medir la brecha digital no necesariamente consiste en utilizar identificadores tecnológicos, sino que se debe considerar la influencia de factores: sociales, económicos, políticos.

La penetración de Internet en los hogares lo ha convertido en un medio masivo de comunicación, con la característica que puede adquirir nuevas formas de comunicación (Ayala, 2007).

2.1.2.1 Internet y el logro académico.

Internet es una poderosa herramienta para la comunicación, pero su uso va más allá, es una oportunidad para ser aprovechada por la educación y transformar la forma de aprender de muchas personas en el mundo.

Para lograr utilizar Internet en el campo educativo es importante tener en cuenta:

- Logar cierto conocimiento sobre el uso de Internet.
- Utilizar Internet como medio para responderse a preguntas y descubrir conocimiento, tener sentido de investigación y crítica (evitar distracciones).
- Diferenciar en el uso de Internet respecto al nivel socioeconómico de las personas. Alguien con mejor nivel cultural, utilizará mejor Internet o aprovechará todo su potencial (Ayala, 2007).
- Descubrir y potenciar habilidades en el uso de Internet para crear, colaborar, compartir, analizar y potenciar como herramienta de aprendizaje e innovación.
- Aprovechar el potencial para aprender desde la libertad, reduciendo barreras de fronteras, tiempo, espacio. Intentar acceder a recursos de calidad desde el e-learning como los MOOC, Wikis, Cursos en Línea, Bibliotecas virtuales, etc.

- Motivar a los estudiantes en utilizar técnicas de Investigación que permitan fomentarles el uso de Internet con un sentido responsable para adquirir conocimiento, lograr un mejor desempeño académico y no como copia o descarga de información. En palabras de (Ayala, 2007) es tomar importancia la “calidad de uso” de Internet, usos que vayan más allá de las satisfacciones de las necesidades que motivan el uso de Internet, tales como el logro educativo, mejora salarial, aumento de la calidad de vida, etc.

Para (Ayala, 2007) el logro académico se entiende como el nivel de cada persona para apropiarse de diferentes tipos de conocimiento de manera que se pueda tener más y mejores formas de satisfacer necesidades. El logro académico va más allá del resultado que se pueda obtener en una evaluación.

2.1.2.2 Internet y las actividades de entretenimiento.

Internet para los jóvenes se ha convertido en un medio de entretenimiento, si bien es cierto muchas personas utilizan Internet como fuente de información, comunicación y estudio, otros la utilizan con fines sociales y de entretenimiento.

Nuestra sociedad está enlazada a la red, como se expresó en otro apartado, somos seres sociales, y esto se produce gracias al estar conectados por una serie de dispositivos que producen, reciben, procesan, transmiten y almacenan información en diferentes condiciones de distancia, tiempo, calidad y cantidad (Parra Castrillón, 2010).

Por esta razón, los jóvenes universitarios no son punto aparte del uso de Internet para actividades de entretenimiento. Los jóvenes en la actualidad buscan en Internet el medio que les permita seguirse relacionando con la sociedad y con ellos mismos, con sus pares.

Internet permite unirse en sociedad, las personas pueden en ejercicio de su libertad ir a lugares físicos o virtuales contruidos para el ocio, la diversión o el juego; para el chantaje o la estafa; para la academia, o la ciencia; para aprender o para comunicarse.

Según lo anterior, las posibilidades de conductas perniciosas no son exclusividad de Internet, sino alternativas del propio mundo (Parra Castrillón, 2010).

En los estudiantes, especialmente los jóvenes, existe tendencia al uso de Internet para la interacción, y en cierto grado una menor tendencia para la búsqueda de la información. Este hecho nos refleja el cambio cultural en el uso de Internet de los más jóvenes para interactuar de manera online. Los estudiantes mayores, por así decirlo, en cambio tienen un aprendizaje clásico y unidireccional (Castaño M, 2011).

2.1.2.3 El nivel de ingresos y el uso de Internet.

Para Bozionelos Nikos citado en (Castaño M, 2011) define que el estatus socioeconómico alto está correlacionado con una mayor experiencia de uso y por tanto con mayores habilidades. Esto significa que el una persona con un nivel económico alto, por su acceso a la educación bien podría obtener mayor acceso y disponibilidad para comprar equipos o dispositivos con acceso a Internet.

Para (Torres, Juan Carlos; Infante, Alfonso, 2011) el nivel de ingresos de la familia de un estudiante incide en los usos e intensidad de uso de las herramientas de Internet, por lo que existe una diferenciación o brecha que se ajusta a la realidad socioeconómica.

Podemos analizar que estas diferencias están marcadas por los costos de uso de Internet, uso de Internet (juegos en línea, televisión, videos y música).

2.1.3 Resultados de medición de la brecha digital a nivel mundial.

Antes de la era de la Internet, la medición del uso telefónico se realizaba a través del *indicador de densidad telefónica*, teledensidad, que consiste en identificar al número de abonados de telefonía fija por cada cien habitantes.

- **IDH (Índice de Desarrollo Humano)**, del *Programa de las Naciones Unidas para el Desarrollo* (www.undp.org), calculado a partir de los siguientes indicadores: Esperanza de vida al nacer, Tasa de alfabetización en adultos, Matrícula escolar, PIB (producto interno bruto) per cápita.
Ecuador en el año 2013, se ubicó en el puesto # 98, con un índice de 0,711 en el que presenta una mejora en relación al 2012, que obtuvo 0,708 (Programa de las Naciones Unidas para el Desarrollo, 2014).
- **IDT (Índice de Desarrollo de las TIC)**, de la *Unión Internacional de Telecomunicaciones* UIT (www.itu.int), que es un índice compuesto que combina 11 indicadores en una medida de referencia que sirve para supervisar y comparar la evolución de las tecnologías de la información y la comunicación TIC entre los países (División de Datos y Estadísticas de TIC de la UIT, 2014).
Este indicador, además, tiene como objetivos principales medir:
 - El nivel y la evolución tecnológica del desarrollo de las TIC.
 - Los progresos de desarrollo de las TIC.
 - La brecha digital.
 - Las posibilidades de desarrollo de las TIC.

Nuestro país en el año 2013, se ubicó en el puesto No. 88, con un IDT de 4,56, a diferencia del 2012 con un IDT de 4,28 (División de Datos y Estadísticas de TIC de la UIT, 2014).

Cuadro 2.1: Índice de Desarrollo de las TIC (IDT), 2012 y 2013

Economía	Puesto 2013	IDT 2013	Puesto 2012	IDT 2012
Dinamarca	1	8,86	2	8,78
Dinamarca	2	8,85	1	8,81
Suecia	3	8,67	3	8,68
Islandia	4	8,64	4	8,58
Reino Unido	5	8,50	7	8,28
Noruega	6	8,39	6	8,35
Países Bajos	7	8,38	5	8,36
Finlandia	8	8,31	8	8,27
Hong Kong, China	9	8,28	11	8,08
Luxemburgo	10	8,26	9	8,19
Japón	11	8,22	10	8,15
Australia	12	8,18	12	8,03
Suiza	13	8,11	13	7,94
Estados Unidos	14	8,02	14	7,90
Mónaco	15	7,93	17	7,72
Singapur	16	7,90	15	7,85
Alemania	17	7,90	18	7,72
Francia	18	7,87	16	7,73
Nueva Zelanda	19	7,82	19	7,62
Andorra	20	7,73	24	7,41
Estonia	21	7,68	21	7,54
Macao, China	22	7,66	20	7,59
Canadá	23	7,62	25	7,37
Austria	24	7,62	23	7,46
Bélgica	25	7,57	26	7,33
Irlanda	26	7,57	22	7,48
Bahrein	27	7,40	28	7,22
España	28	7,38	29	7,14
Israel	29	7,29	27	7,25
Malta	30	7,25	30	7,08
Eslovenia	31	7,13	31	6,96
Emiratos Árabes Unidos	32	7,03	46	6,27
Letonia	33	7,03	33	6,84
Qatar	34	7,01	42	6,46
Barbados	35	6,95	32	6,87
Italia	36	6,94	36	6,66
Croacia	37	6,90	34	6,70
Bielorrusia	38	6,89	43	6,45
Grecia	39	6,85	35	6,70
Letonia	40	6,74	40	6,50
República Checa	41	6,72	38	6,57
Federación de Rusia	42	6,70	41	6,46
Portugal	43	6,67	39	6,57
Polonia	44	6,60	37	6,63
Eslovacaquia	45	6,58	45	6,30
Hungría	46	6,52	44	6,35
Arabia Saudita	47	6,36	50	6,01
Uruguay	48	6,32	51	5,92
Bulgaria	49	6,31	47	6,12
Serbia	50	6,24	49	6,07
Chipre	51	6,11	48	6,09
Omán	52	6,10	61	5,43
Kazajistán	53	6,08	62	5,80
Saint Kitts y Nevis	54	6,01	52	5,89
Costa Rica	55	5,92	55	5,64
Chile	56	5,92	54	5,68
Antigua y Barbuda	57	5,89	59	5,49
Rumania	58	5,83	58	5,52
Argentina	59	5,80	56	5,58
Laos R. P. y Macedonia	60	5,77	63	5,42
Moldavia	61	5,72	60	5,44
Líbano	62	5,71	64	5,32
Montenegro	63	5,67	57	5,52
Azerbaiján	64	5,65	65	5,22
Brasil	65	5,50	67	5,16
Brunei Darussalam	66	5,43	63	5,36
Trinidad y Tobago	67	5,29	70	4,99
Turquía	68	5,29	68	5,12
Bosnia y Herzegovina	69	5,23	74	4,89
Mauricio	70	5,22	72	4,96
Malasia	71	5,20	66	5,18
San Vicente y las Granadinas	72	5,17	69	5,04
Ucrania	73	5,15	71	4,97
Armenia	74	5,08	73	4,89
Seychelles	75	4,97	76	4,70
Granada	76	4,96	75	4,83
Colombia	77	4,95	80	4,61
Georgia	78	4,86	83	4,48
Santa Lucía	79	4,81	79	4,66
Venezuela	80	4,81	78	4,68
Tailandia	81	4,76	91	4,09
Panamá	82	4,75	77	4,69
Dominica	83	4,72	81	4,58

Economía	Puesto 2013	IDT 2013	Puesto 2012	IDT 2012
Albania	84	4,72	85	4,42
Maldivas	85	4,71	82	4,50
China	86	4,64	86	4,29
Jordania	87	4,62	84	4,48
Ecuador	88	4,56	88	4,28
Egipto	89	4,45	87	4,28
Sudafricana (Rep.)	90	4,42	89	4,19
Fiji	91	4,40	103	3,90
Mongolia	92	4,32	90	4,19
Cabo Verde	93	4,30	104	3,86
Irán (R. del)	94	4,29	97	4,02
México	95	4,29	94	4,07
Marruecos	96	4,27	92	4,09
Jamaica	97	4,26	96	4,01
Suriname	98	4,26	93	4,08
Túnez	99	4,23	96	4,07
Palestina	100	4,16	95	4,07
Viet Nam	101	4,09	99	3,94
Dominicana (Rep.)	102	4,06	105	3,78
Filipinas	103	4,02	102	3,91
Botswana	104	4,01	100	3,94
Perú	105	4,00	101	3,92
Indonesia	106	3,83	106	3,70
Bolivia	107	3,78	109	3,52
Kingstón	108	3,78	107	3,59
Paraguay	109	3,71	108	3,56
El Salvador	110	3,61	110	3,47
Guyana	111	3,46	111	3,44
Sri Lanka	112	3,46	112	3,39
Ghana	113	3,46	115	3,29
Algeria	114	3,42	114	3,30
Uzbekistán	115	3,40	116	3,27
Sri Lanka	116	3,36	113	3,31
Namibia	117	3,24	118	3,08
Guatemala	118	3,20	117	3,11
Honduras	119	3,18	119	3,01
Nicaragua	120	2,96	120	2,78
Zimbabue	121	2,89	123	2,68
Sudán	122	2,88	121	2,69
Bhután	123	2,85	126	2,58
Kenya	124	2,79	124	2,62
Cuba	125	2,77	122	2,69
Guadalupe	126	2,66	125	2,61
Camboya	127	2,61	127	2,54
Suazilandia	128	2,60	128	2,43
India	129	2,53	129	2,42
Senegal	130	2,46	133	2,20
Nepal	131	2,37	134	2,20
Lesotho	132	2,36	131	2,22
Nigeria	133	2,35	135	2,14
Laos (R.D.P.)	134	2,35	130	2,25
Gambia	135	2,31	136	2,12
Salomón (Islas)	136	2,29	132	2,22
Congo (Rep.)	137	2,24	137	2,09
Yemen	138	2,18	138	2,07
Angola	139	2,17	139	2,06
Camerún	140	2,10	142	1,98
Djibouti	141	2,08	140	2,01
Pakistán	142	2,05	141	2,01
Malí	143	2,04	147	1,86
Zambia	144	2,02	143	1,97
Bangladesh	145	1,97	146	1,90
Uganda	146	1,94	144	1,90
Mauritania	147	1,91	145	1,90
Rwanda	148	1,86	151	1,74
Benin	149	1,84	149	1,75
Myanmar	150	1,82	148	1,75
Côte d'Ivoire	151	1,80	150	1,74
Tanzania	152	1,76	152	1,72
Liberia	153	1,70	154	1,57
Guinea-Bissau	154	1,67	153	1,60
Afghanistan	155	1,67	155	1,57
Burkina Faso	156	1,56	160	1,35
Congo (Rep. Dem. del)	157	1,56	157	1,47
Malawi	158	1,52	156	1,50
Mozambique	159	1,52	159	1,40
Madagascar	160	1,42	158	1,43
Guinea	161	1,42	161	1,31
Etiopía	162	1,31	162	1,24
Eritrea	163	1,20	163	1,18
Chad	164	1,11	164	1,09
Niger	165	1,03	165	0,97
Centrosurfricana (Rep.)	166	0,96	166	0,93

Fuente: UIT.

Figura 1: Índice de Desarrollo de las TIC 2012 y 2013

Fuente: Informe sobre medición de la Sociedad de Información 2014

Elaborado por: Unión Internacional de Telecomunicaciones (ITU).

- **NRI (Networked Readiness Index)**, del Foro Económico Mundial, es un indicador compuesto que mide la habilidad de una economía para apalancar sus avances en las TICs, en beneficio de su competitividad y el Buen Vivir de sus ciudadanos. Según este indicador, que forma parte del estudio *Global Information Technology Report (GITR)*, nuestro País se ubica en el lugar No. 82,

con un índice de 3,9 (Ministerio de Telecomunicaciones y Sociedad de la Información, 2014).

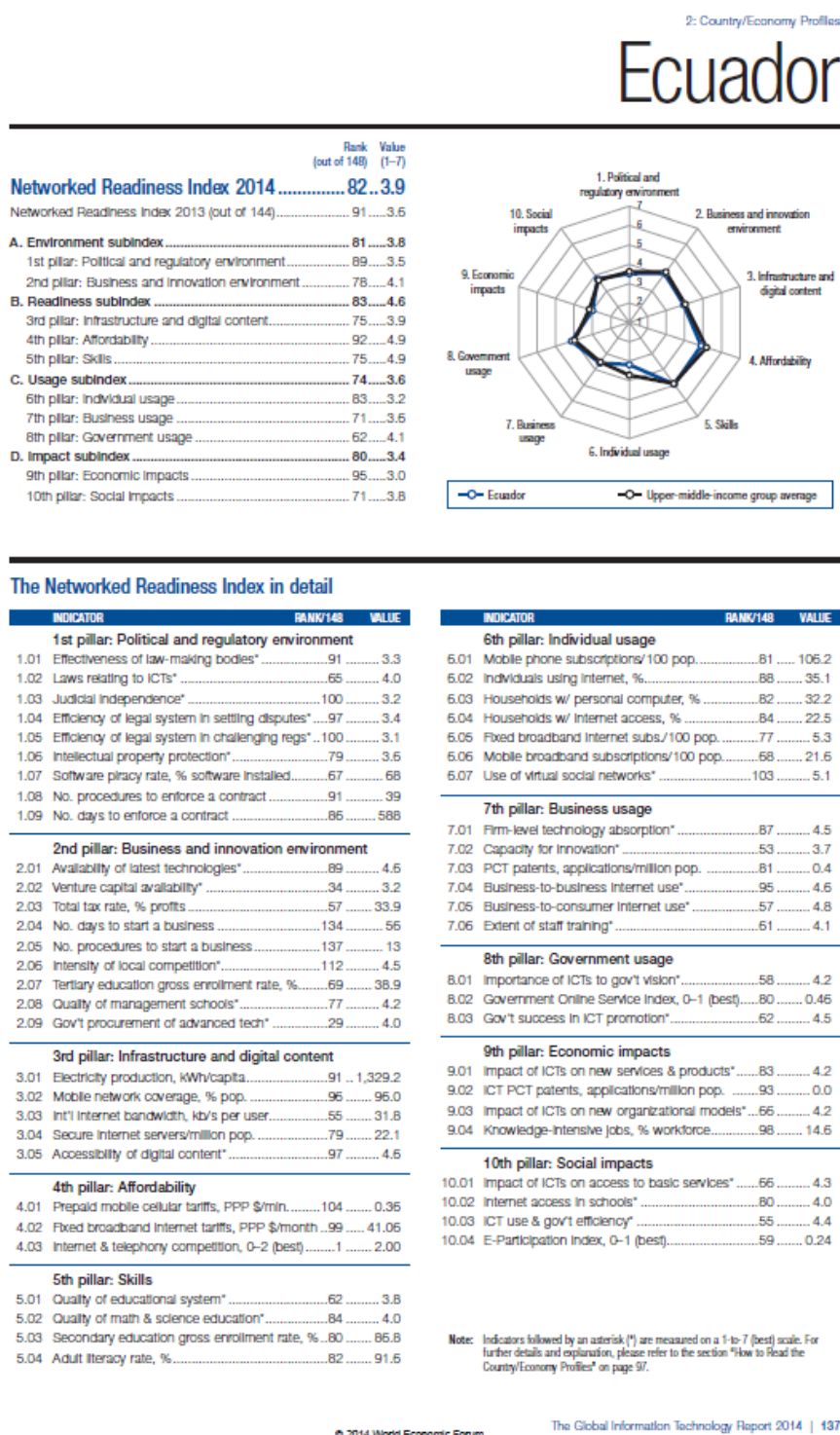


Figura 2: Networked Readiness Index Ecuador 2014
Fuente: The Global Information Technology Report 2014
Elaborado por: Foro Económico Mundial

2.2 Acceso a las Tecnologías de la Información y Comunicaciones en el Ecuador.

Según (INEC - Instituto Nacional de Estadística y Censos, 2012) las Tecnologías de la Información y la Comunicación en nuestro País reflejan que al menos 6,8 millones de personas tienen un celular activado, 800 mil personas tienen un teléfono inteligente, 600 mil personas usan su teléfono para acceder a redes sociales (Facebook, Twitter, YouTube) de los cuales la mayoría reside en el área urbana, es soltero, cuenta con instrucción superior universitaria, tiene ocupación plena y es empleado privado. La Provincia de Guayas, registra el mayor número de personas que tiene un teléfono celular inteligente 20,8%.

Además para (INEC - Instituto Nacional de Estadística y Censos, 2013), las Estadísticas Sociales en el uso de las Tecnologías de la Información y Comunicaciones en Ecuador, basadas en los resultados de la Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDUR 2010 - 2013), revela que: el 18,1% de los hogares ecuatorianos tiene al menos un computador portátil, mientras que el 27,5% tiene un computador de escritorio, el 86,4% posee al menos un teléfono celular, el 28,3% tiene acceso a Internet en el hogar de los cuales un 31,4% accede vía inalámbrica, 24,9% a través de un sistema de banda ancha y un 43,7% vía módem o teléfono, el 43,6% utilizaron una computadora, el grupo etario con mayor número de personas que utilizaron computadora es el que está entre 16 a 24 con el 67,8%, seguidos de 5 a 15 años con 58,3%, el 40,4% de la población accedió a Internet en el 2013, de los que el 45,1% lo hizo desde su hogar, 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 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%. El 16,9% de las personas que posee un celular tiene un teléfono inteligente (smartphone). En el 2013, el 20,0% de las personas en el Ecuador son analfabetas digitales. Según (Villacís & Jaramillo, 2014) la mayoría era mujeres (50,4%) y la edad fue un factor predominante. Apenas el 18,9% de las personas de entre 55 y 64 años había usado una computadora. Frente al 67,8% de la población de entre 16 y 24 años que sí lo hizo.

2.3 Minería de datos.

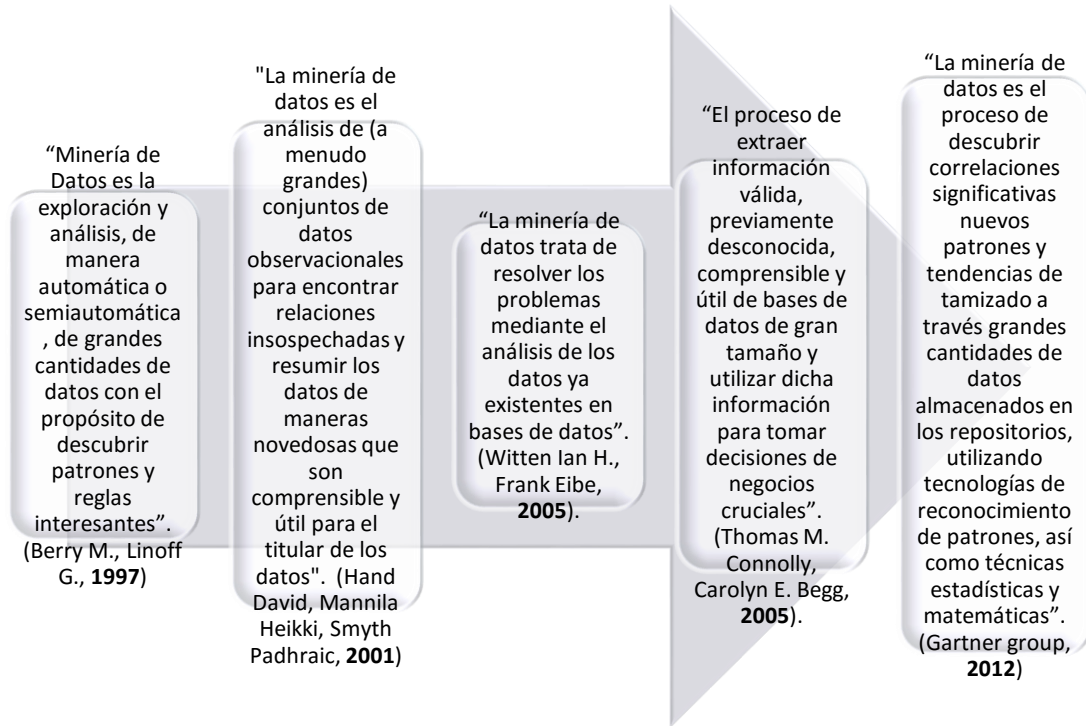


Figura 3: Evolución del Concepto de Minería de Datos

Fuente: Elaboración del autor

2.3.1 Operaciones y Técnicas de minería de datos.

(Pérez López & Santín González, 2007), minería de datos distingue técnicas predictivas y descriptivas.

Las *técnicas predictivas* (supervisadas) especifican el modelo de los datos en base a un conocimiento previo, las variables se clasifican en dependientes e independientes. Las *técnicas descriptivas* (no supervisadas), las variables tienen el mismo status, ver figura 4.

Estas técnicas permiten obtener conocimiento a partir de los datos.

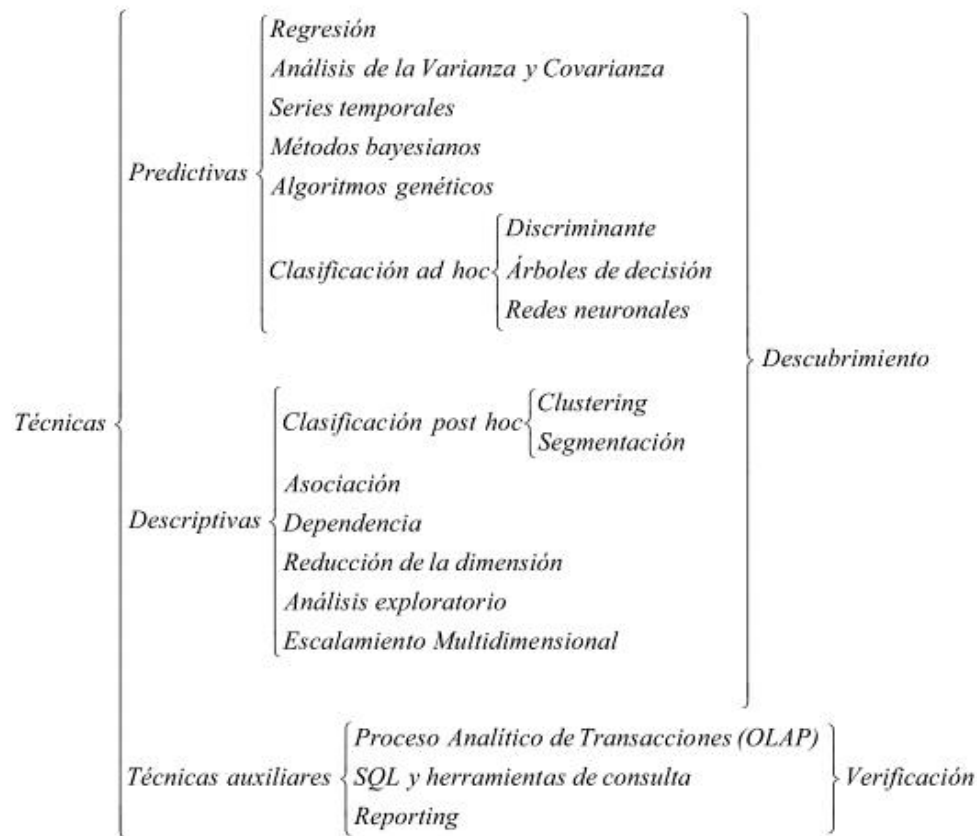


Figura 4: Técnicas de Minería de Datos
Fuente: (Pérez López & Santín González, 2007)

Hay cuatro operaciones principales (y técnicas asociadas) asociadas con las técnicas de minería de datos, que son (Connolly & Begg, 2005):

Tabla 1: “Operaciones de minería de datos y sus técnicas asociadas”

Operaciones	Técnicas de minería de datos
Modelado predictivo:	Clasificación. Predicción de valores.
Segmentación de la base de datos:	Agrupaciones demográficas. Agrupaciones neuronales.
Análisis de enlaces:	Descubrimiento de asociaciones. Descubrimiento de patrones secuenciales. Descubrimiento de similitudes en secuencias temporales.
Detección de desviaciones:	Estadística. Visualización.

Fuente: (Connolly & Begg, 2005)

2.3.2 Aplicaciones de la minería de datos.

Existen numerosas áreas donde la minería de datos se puede aplicar, prácticamente en todas las bases de datos de actividades humanas que generen datos e información. Entre muchos usos, principalmente podemos destacar las siguientes aplicaciones de la minería de datos (Riquelme, Ruiz, & Gilbert, 2006):

- **Comercio y banca:** segmentación de clientes, previsión de ventas, análisis de riesgo y fraude.
- **Medicina y farmacia:** diagnóstico de enfermedades y la efectividad de los tratamientos.
- **Seguridad y detección de fraude:** reconocimiento facial, identificaciones biométricas, accesos a redes no permitidos, etc.
- **Recuperación de información no numérica:** minería de texto, minería WEB, búsqueda e información de imagen, video, voz y texto de bases de datos multimedia.
- **Astronomía:** identificación de nuevas galaxias y estrellas.
- **Geología, minería, agricultura y pesca:** identificación de áreas para uso de distintos cultivos o pesca o explotación minera en bases de datos de imágenes satelitales.
- **Ciencias ambientales:** identificación de modelos de funcionamiento de ecosistemas naturales y/o artificiales (p. ejm. Plantas depuradoras de aguas residuales) para mejorar su observación, gestión y/o control.
- **Ciencias sociales:** Estudio de los flujos de opinión pública. Planificación de ciudades: identificar barrios en conflicto en función de valores sociodemográficos.
- **Administración pública y servicios:** Análisis y control de tráfico de vehículos, censos poblacionales, económicos y agropecuarios (Galvis & Martínez, 2004).
- **Educación:** predicción del comportamiento estudiantil, análisis de los patrones de deserción y abandono estudiantil, tendencias en la elección de una carrera o especialización, etc.

2.3.3 Minería de Datos en la Educación (MDE), Educational Data Mining (EDM).

La International Educational Data Mining Society, define a la MDE como: *“Una disciplina emergente, concerniente con el desarrollo de métodos para explorar los únicos tipos de datos que provienen de temas educacionales, y usa éstos métodos para entender mejor a los estudiantes y a los temas concernientes con su aprendizaje”* (International Educational Data Mining Society, 2012).

MDE es la aplicación de las técnicas de minería de datos para los datos extraídos en el campo educativo, y además, su objetivo es analizarlos en orden a resolver cuestiones investigativas educacionales (Barnes, Desmarais, Romero, & Ventura, 2009).

Según (Calders & Pechenizkiy, 2012) MDE es una ciencia de aprendizaje, así como un área rica de aplicación para la minería de datos, debido a la creciente disponibilidad de datos educativos.

La MDE contribuye al estudio y análisis del aprendizaje de los estudiantes, así como el desarrollo de métodos para explorar los tipos únicos de datos obtenidos en el ámbito educativo y, con estos métodos entenderlos mejor y el contexto en los que aprenden. Además permite la toma de decisiones basada en la mejora de la práctica educativa actual y material de aprendizaje, cursos, oferta educativa, investigación en materia docente y administrativa de centros educativos.

“MDE utiliza los repositorios de datos para entender mejor a los estudiantes y el aprendizaje” (Romero & Ventura, Educational Data Mining: A review of the State-of-the-Art, 2010).

2.3.3.1 Técnicas aplicadas en minería de datos en la educación.

La MDE convierte los datos en bruto procedentes de sistemas educativos en información útil que podría tener un gran impacto en la investigación y la práctica educativa y éste proceso no difiere mucho de otras aplicaciones de la minería de datos tales como negocios, genética, medicina, etc., porque sigue los mismos pasos del proceso de

minería de datos: preprocesamiento, minería, postprocesamiento (Romero, De Bra, & Ventura, "Knowledge discovery with genetic programming for providing feedback to courseware autor"., 2004).

Ryan Baker (2010), clasifica las técnicas de minería de datos para la educación como sigue:

- Predicción:
 - Clasificación.
 - Regresión.
 - Estimación de la densidad.
- Clustering.
- Minería de relaciones:
 - Regla de asociación
 - Minería correlacional.
 - Minería de secuencia de patrones.
 - Minería de datos causal.
- Destilación de datos para el juicio humano.
- Descubrimiento con modelos.

Desde un práctico punto de visto, la MDE permite descubrir nuevo conocimiento a partir de los datos basados en el comportamiento de los estudiantes de manera que permita ayudar a validar/evaluar sistemas educacionales, para potenciar y mejorar algunos aspectos de la calidad de la educación y sentar las bases para un mejor efectivo proceso de aprendizaje (Romero, De Bra, & Ventura, "Knowledge discovery with genetic programming for providing feedback to courseware autor"., 2004).

Hay algunas cuestiones importantes que diferencian la MD de la educación desde cómo es aplicada en otros dominios:

Tabla 2: Diferencias entre Minería de Datos y Minería de Datos en la Educación

Cuestiones	Minería de Datos	Minería de Datos en la Educación
Objetivo:	En cada área es distinto. En negocios el objetivo principal es incrementar los beneficios.	Aplica objetivos de la investigación en procesos como aprendizaje y guía para estudiantes, profundización en entender el fenómeno educativo.
Datos:	Gran cantidad de datos para ser analizados, depurados, limpiados para aplicar el proceso de minería en diferentes aspectos.	En ambientes educativos hay varios tipos diferentes de datos para minería. Datos específicos para el área educacional y su relación con otros datos en múltiples niveles de jerarquía.
Técnicas:	Diferentes técnicas/herramientas de minería pueden ser aplicadas en un sector concreto.	Los datos educacionales y sus problemas tienen algunas características especiales que requieren ser tratados en diferente forma. Técnicas específicas de minería de datos pueden ser utilizadas para problemas educativos específicos.

Fuente: (Romero & Ventura, "Educational Data Mining: a survey from 1995 to 2005", 2007)

MDE involucra diferentes grupos de usuarios o participantes en el proceso de MDE:

La tabla 3, nos muestra los usuarios o participantes y el objetivo de utilizar minería de datos:

Tabla 3: Usuarios de MDE / Beneficiarios

Usuarios / Actores	Objetivos para utilizar minería de datos
Aprendices / Estudiantes / Pupilos	Personalizar e-learning: recomendar actividades para aprendices y recursos, tareas de aprendizaje que pueden mejorar aún más su aprendizaje; sugerir interesantes

	experiencias de aprendizaje a los estudiantes, sugerir rutas y cortas o simples links para seguir, generar objetivos adaptativos, recomendar cursos, discusiones relevantes, libros, etc.
Educadores / Profesores / Instructores / Tutores	Obtener información objetiva acerca de la instrucción; analizar el comportamiento y aprendizaje de estudiantes; detectar qué estudiantes requieren ayuda (recuperación); predecir rendimiento de los estudiantes; clasificar aprendizajes en grupos; encontrar comportamientos regulares o irregulares de los estudiantes; encontrar los errores más comunes; determinar actividades más efectivas; mejorar la adaptación y personalización de cursos, etc.
Curso / Desarrolladores / Investigadores educativos	Evaluar y mantener material didáctico; mejorar el aprendizaje de estudiantes; evaluar la estructura de contenido de curso y su efectividad en el proceso de aprendizaje; construir automáticamente modelos de estudiantes y tutores; comparar técnicas de minería de datos en orden a ser viable recomendar las más útiles para cada tarea; desarrollar herramientas específicas de minería de datos para propósitos educativos; “predecir deserción (mi agregado personal)” etc.
Organizaciones / Proveedores de aprendizaje / Universidades / Compañías privadas de entrenamiento	Mejorar los procesos de decisión en instituciones de educación superior; simplificar eficiencia en la toma de decisiones; lograr objetivos específicos; sugerir algunos cursos que pueden ser valorados por cada clase de aprendices; encontrar la mejor relación costo – efectividad para mejorar la retención (fidelidad) y grados; seleccionar los mejores aplicantes calificados para graduación; ayudar a admitir estudiantes que tendrán éxito en la universidad; “calificar el rendimiento académico de los maestros; predecir los niveles de aceptación, deserción, promoción de estudiantes; identificar las necesidades académicas de un plantel; analizar y tomar decisiones económicas en torno a servicios ofrecidos, etc.” (agregado por el autor).
“Colegios educativos particulares / Instituciones Académicas de educación continua (agregado por el autor)”	
Administradores / Supervisores de distritos escolares	Desarrollar la mejor vía para organizar los recursos institucionales (humanos y materiales) y su oferta

Administradores de redes / educativa; utilizar los recursos disponibles
 Administradores de sistemas eficientemente; mejorar la oferta educativa, determinar el
 “Rectores / Vicerrectores / acercamiento de la educación a distancia; evaluar el
 Rectores Administrativos y currículum de un maestro; establecer parámetros para
 Pedagógicos / Coordinadores de mejorar la eficiencia y adaptabilidad de un sitio WEB a
 Nivel, etc. (agregado por el autor)”. los usuarios optimizando servicios, red de distribución
 tráfico; “predecir niveles de aceptación, deserción,
 abandono escolar (agregado por el autor)”.

Fuente: (Romero & Ventura, "Educational Data Mining: A Review of the State-of-the-Art", 2010)

2.3.4 Procedimiento para descubrir la información oculta en bases de datos o KDD.

La minería de Datos, es parte del proceso de descubrimiento de conocimiento a partir de los datos.

Según (Molina López & García Herrero, 2006) KDD es el proceso que se encarga de la preparación, extracción de información e interpretación de resultados obtenidos.

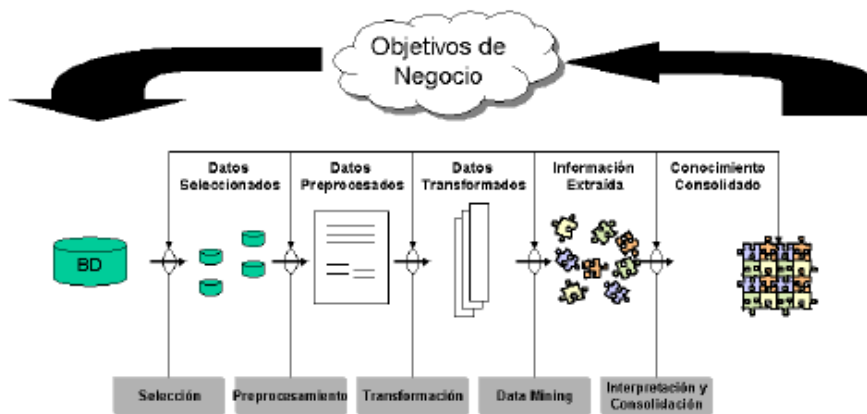


Figura 5: Metodología para el descubrimiento de conocimiento en bases de datos

Fuente: (Molina López & García Herrero, 2006)

KDD se ha definido como “el proceso no trivial de identificación en los datos de patrones válidos, nuevos, potencialmente útiles, y finalmente comprensibles” (U. Fayyad, 1996).

Para (Hernández Orallo, 2006) el conocimiento a partir de los datos se obtiene luego de un proceso interactivo, bajo el criterio de un experto en toma de decisiones, consta de varias fases que como la preparación de los datos (selección, limpieza y transformación), exploración y auditoría, aplicación de la minería de datos (extracción de información a partir de modelos y algoritmos), interpretación y consolidación (Pérez López & Santín González, 2007).

Por lo tanto, podemos definir que el proceso de KDD, consta de los siguientes pasos:

- Limpieza e integración.
- Selección y transformación.
- Minería de Datos.
- Evaluación y presentación.
- Conocimiento.

2.3.4.1 Limpieza e integración.

Se limpian los datos inconsistentes, atípicos y con ruido. Se combinan fuentes de datos (si fuera el caso) (Han, Kamber, & Pei, 2012).

2.3.4.2 Selección y transformación.

Identificar y seleccionar las variables relevantes de los datos (Pérez López & Santín González, 2007). Transformar y consolidar los datos en formas apropiadas para la Minería de Datos (Han, Kamber, & Pei, 2012).

2.3.4.3 Minería de Datos.

Proceso esencial en el que aplicando algoritmos y métodos inteligentes se extrae patrones de los datos. Minería de Datos es el proceso de descubrimiento de patrones interesantes y conocimiento de grandes cantidades de datos. Las fuentes pueden ser bases de datos, Internet, o datos generados dinámicamente desde transmisiones (Han, Kamber, & Pei, 2012).

2.3.4.4 Evaluación y presentación.

La evaluación permite identificar la veracidad de los datos obtenidos que representan el conocimiento, a través de los modelos (Pérez López & Santín González, 2007). Para definir la veracidad de los datos es importante que tenga las siguientes características:

- Entendibles fácilmente por los humanos.
- Validar o testear los datos con grados de certeza.
- Que sean potencialmente útiles.
- Novedosos (Han, Kamber, & Pei, 2012).

2.3.4.5 Presentación del Conocimiento.

El conocimiento generado es presentado a los usuarios a través de técnicas de visualización y simulación, gráficas, reportes, informes o conclusiones (Han, Kamber, & Pei, 2012).

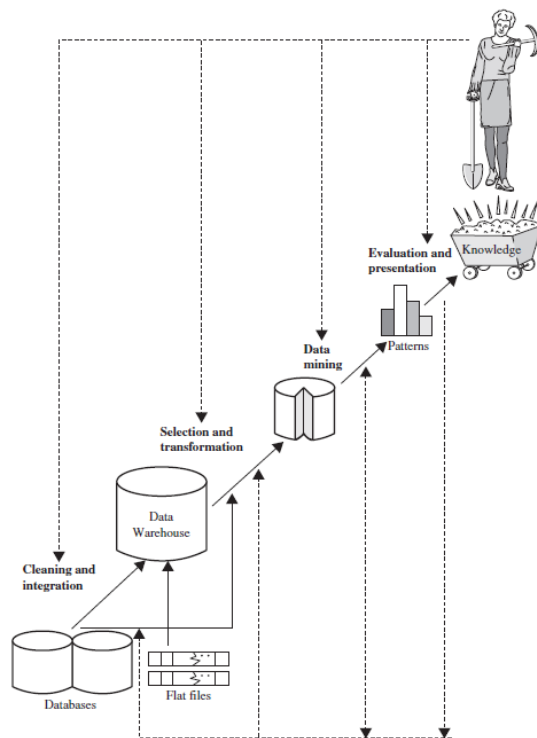


Figura 6: Proceso de obtención del conocimiento KDD
Fuente: (Han, Kamber, & Pei, 2012)

CAPÍTULO III: METODOLOGÍA DE INVESTIGACIÓN

3.1 Población y Tamaño de la muestra.

La población del presente trabajo de investigación, corresponde a los estudiantes de la Escuela Superior Politécnica del Litoral, perteneciente a 37 carreras.

En el período lectivo 2014 – 2015, la ESPOL contó con 10778 estudiantes matriculados.

Para calcular el tamaño de la muestra cuando el universo es finito utilizamos la fórmula para el cálculo de la muestra de poblaciones finitas (Arnal, del Rincón, & Latorre, 1992):

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

Donde:

- n** Tamaño de la muestra
- Z** Valor de confianza del 95% = 1,96.
- P Q** Proporción esperada (5%).
- N** Tamaño de la población
- d** Precisión de la investigación (5%).

Reemplazando tenemos:

$$n = \frac{1,96^2 \times 0.5 \times 0.5 \times 10778}{0,05^2(10778 - 1) + 1,96^2 \times 0.5 \times 0.5}$$

$$n = \frac{10351,1912}{27,9029}$$

$$n = 370$$

Si bien es cierto, el tamaño de la muestra corresponde a 370 estudiantes, para maximizar el rendimiento de la presente investigación, esta muestra se amplía a 490 estudiantes.

3.2 Procedimiento para descubrir el conocimiento en el Proyecto de Investigación (KDD).

En este apartado se aplican los componentes que forman el proceso para descubrir el conocimiento en el Proyecto de Investigación.

3.2.1 Recolección de Información.

El instrumento para recolectar la información es la encuesta correspondiente al Proyecto “La Educación virtual en el Ecuador” y está basada en los proyectos Internet Cataluña (PIC), Proyecto Digital Literacy in Higher Education (DLINHE) y a los lineamientos del autor del Trabajo de Tesis Doctoral: *“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”* (Torres, 2012) (Anexo 1).

3.2.1.1 Descripción del contenido del instrumento de recolección de información.

Para la medición de las variables se aplican escalas de tipo Likert², clasificados de la siguiente manera:

Variables con información del estudiante y variables sociodemográficas

Universidad, Carrera del estudiante, edad, género, ingresos mensuales de la familia.

Variables que miden el nivel de conocimiento y el uso de Internet

² Rensis Likert (1932) –Ítems presentados en forma de juicios para medir la reacción de los encuestados.

Lugar de conexión, número de días a la semana que se conecta a Internet, nivel de conocimientos en el manejo de Internet, horas de conexión y años de experiencia como usuario de Internet.

Variables que miden el uso de Internet con fines académicos

Número de veces a la semana que ingresa a la plataforma virtual de la Universidad, consultas realizadas a los profesores, consultas a compañeros, recursos educativos descargados desde la plataforma virtual cada mes, videos vistos en YouTube con fines académicos, participación mensual en foros virtuales, post o tuits sobre temas académicos, búsqueda de información académica en Internet cada vez, uso de la biblioteca virtual cada mes. Variables de tipo nominal.

Variables que miden el uso de Internet con fines sociales, entretenimiento y diversión

Número de horas semanales para: chat por diversión, redes sociales, juegos en línea, descargas de música, videos y programas, videos para entretenimiento vistos en YouTube, variables de tipo nominal.

Preguntas referentes al uso de un blog, cuenta en YouTube y del.icio.us de tipo dicotómicas.

Variables que informan sobre el rendimiento académico

Asignaturas matriculadas, asignaturas aprobadas.

La encuesta tiene un total de 17 ítems, y estos a su vez contienen varios subítems que engloban los elementos de estudio en la presente tesis. Este instrumento se aplicó de manera presencial en la ESPOLE y el ingreso de la información se realizó en forma virtual, utilizando el aplicativo de Survey Monkey³.

³ <https://es.surveymonkey.com/mp/aboutus/>

3.2.2 Procedimiento para la fase de Selección y Transformación de los datos.

Luego de realizar la encuesta de manera presencial, se procedió a subir los resultados de la misma a través del aplicativo en la WEB Survey Monkey, a través del enlace proporcionado por el Director del Proyecto⁴.

Se analiza la calidad de los datos seleccionados de la encuesta. La calidad de los datos se compone de tres elementos: *exactitud*, *integridad* y *consistencia*. (Han, Kamber, & Pei, 2012). En un proceso de KDD, se deben detectar los datos anómalos, es por ello que se debe eliminar los valores en blanco, ruido, datos duplicados, valores atípicos.

Se utilizó el Software IBMS SPSS⁵ (Statistical Package for the Social Sciences) versión 19, el archivo con los datos fue proporcionado por nuestro Director del Proyecto.

Se procedió a detectar, identificar y corregir:

- Nombres de las carreras ofertadas por la ESPOL.
- Detección de datos atípicos, o fuera del rango.
- Eliminación de registros con datos incompletos o ambiguos.
- Eliminación de registros en blanco.
- Corrección de valores atípicos, a través de los estadísticos mediana y moda.
- Para el rendimiento académico, se creó una nueva variable, la variable "Rendimiento" que es la diferencia entre las variables de: Asignaturas Aprobadas y Asignaturas Matriculadas.

Una vez construido el set datos se procedió a seleccionar las variables a utilizar para realizar la exploración y el experimento con los datos para llegar a la fase de minería de datos, de manera que nos permita tener los perfiles de análisis en lo correspondiente a los usos de Internet, tanto para actividades académicas, como para entretenimiento.

Estas variables corresponden a los grupos de:

- Ingresos mensuales: Variable: *Ingresos*.

⁴ Enlace para subir los datos de la encuesta en la ESPOL

<https://es.surveymonkey.com/r/?sm=WUR%2fQLmAEH6wg1zELq2spevF2Qv4yadGrBv9fQWOTT0%3d>

⁵ <http://www-01.ibm.com/software/analytics/spss/>

- Usos de Internet con fines académicos. Variables: *A_IngresoPlataforma*, *A_ConsultasProfesor*, *A_ConsultasEstudiantes*, *A_RecursosEducativos*, *A_VideosAcadémicos*, *A_ForosVirtuales*, *A_PostsAcadémicos*, *A_HorasChatAcadémico*, *A_HorasBúsqueda*, *A_BibliotecaVirtual*.
- Uso de Internet con fines de entretenimiento. Variables: *E_HorasChatPersonal*, *E_RedesSociales*, *E_JuegosLínea*, *E_DescargaMúsica*, *E_Videos*.
- Variables que miden el rendimiento: *SA_AsignaturasMatriculadas*, *SA_AsignaturasAprobadas*, *Rendimiento*.

3.2.3 Minería de Datos.

Esta fase corresponde a los análisis estadísticos realizados. Estadística estudia la colección, análisis, interpretación o explicación, y presentación de los datos. La minería de datos tiene una conexión inherente con la estadística (Han, Kamber, & Pei, 2012). Un modelo estadístico es un set de funciones matemáticas que describen el comportamiento de los objetos en una clase. Los modelos estadísticos pueden ser la salida o el resultado de una técnica de la minería de datos.

3.2.3.1 Análisis de Correspondencia Simple ACS.

El Análisis de Correspondencia es una técnica descriptiva o exploratoria cuyo objetivo es resumir una gran cantidad de datos en un número reducido de dimensiones (de la Fuente Fernández, 2011).

El análisis de Correspondencia Simple, es la técnica en la cual se evalúa la relación existente en una tabla de contingencia de dos variables ordinales o categóricas. Se orienta a casos en los cuales una variable representa ítems y cualidades de un grupo de individuos, organizados en categorías (Salazar Paredes, 2012).

El Análisis de Correspondencia tiene dos objetivos básicos:

- Medir la asociación de una fila o columna.
- Estudiar la relación (si existe) entre filas y columnas (de la Fuente Fernández, 2011).

3.2.3.2 χ^2 - *chi cuadrado*.

El contraste de interdependencia entre las variables fila y columna en una tabla de contingencia I x J, se realiza mediante la utilización del estadístico χ^2 :

$$\chi^2 = \sum \frac{(\text{frecuencias observadas} - \text{frecuencias esperadas})^2}{\text{frecuencias esperadas}}$$

que en la hipótesis de interdependencia sigue una distribución de χ^2 con $(I - 1) \times (J - 1)$ grados de libertad. De acuerdo con la notación, la frecuencia esperada en cada celda de la fila i , suponiendo independencia de filas y columnas, se obtendrá repartiendo el total de la fila nf_i , proporcionalmente a la frecuencia relativa de cada columna, f_j (Peña, 2012).

Por lo tanto el estadístico χ^2 suele escribirse:

$$\chi^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{(nf_{ij} - nf_i \cdot f_j)^2}{nf_{ij} \cdot f_j}$$

Para comprobar el contraste de hipótesis se utiliza el estadístico χ^2 de Pearson, contrastando la hipótesis nula que presupone la independencia de ambas variables.

H_0 : Ambas variables son independientes.

H_1 : Existe una relación de dependencia.

Si (p-valor $\leq 0,05$) se rechaza H_0

Si (p-valor $\geq 0,05$) se acepta H_1 (de la Fuente Fernández, 2011).

3.2.3.3 *Análisis Factorial*.

El Análisis Factorial es una de las técnicas de reducción de datos que sirve para encontrar grupos homogéneos de variables a partir de un conjunto numeroso de ellas

(Ruiz & Pardo, 2001). Estas variables correlacionan entre sí y procurando que, unos grupos sean independientes de otros.

Su propósito es buscar el número mínimo de dimensiones que expliquen el máximo de información contenida en los datos.

Es una técnica de reducción de la dimensión de los datos. Las variables son independientes en el sentido que no existe a priori una dependencia conceptual de una variable sobre otra (Ruiz & Pardo, 2001).

Según (Torres, 2012), para realizar el análisis factorial es necesario identificar 4 elementos:

1. Obtener la matriz de correlaciones a partir de la matriz de datos iniciales. El determinante de esta matriz debe ser bajo y distinto de 0, así se asume que el nivel de correlación de las variables es alto, lo que es un indicador de validez del análisis factorial.
2. El test de esfericidad de Barlett, que permite verificar la hipótesis nula de que la matriz de correlaciones es una matriz de identidad, por ende en las variables no existe una correlación significativa.

Prueba de esfericidad de Barlett:

Si sig (p-valor) < 0.05 aceptamos H_0 , se puede aplicar análisis factorial.

Si sig (p-valor) > 0.05 rechazamos H_0 , no se puede aplicar análisis factorial (Universidad de Alicante, 2011).

3. Prueba de KMO (Kaiser – Meyer – Olkin), que compara los coeficientes de correlación general con los de correlación parcial entre pares de variables. El valor resultante fluctúa entre 0 y 1. Cuanto más cerca de 1 tenga el valor obtenido del KMO, implica que la relación entre variables es alta. (Universidad de Alicante, 2011)
4. Interpretar los factores y determinar el ajuste del modelo.

Se realizaron experimentos con las variables relacionadas con los usos académicos y los usos de entretenimiento, pero se comprueba que el modelo no se ajusta puesto que el valor del determinante es igual a 0.370 en los usos académicos y 0.548 en los usos de entretenimiento; estos valores deben ser bajos y distintos de 0 (Ver anexo 2 y 3).

3.2.3.4 Análisis de conglomerados o análisis clúster.

Es una técnica multivariante que permite agrupar las variables en estudio en un conjunto en función de su parecido o características comunes. Una ventaja del análisis clúster es que permite detectar el número óptimo de grupos y su composición a partir de la similitud que existe entre los casos (Ruiz & Pardo, 2001).

K-medias es un algoritmo de particionamiento, el centro de cada clúster es representado por el valor de la media de los objetos en el clúster.

Entrada:

k : el número de clústers.

D : el set de datos que contiene n objetos.

Salida:

Un set de k clústers.

Procedimiento (algoritmo):

Selección arbitraria de k objetos desde D como inicio de los centros de los grupos.

Repita:

(Re)asignación de cada objeto al clúster al que el objeto sea el más similar, basados en la media de los objetos del clúster.

Actualice las medias de los clústers, calculando la media de los objetos para cada clúster.

Hasta: que no hayan cambios (Han, Kamber, & Pei, 2012).

Luego de los experimentos y análisis realizados con el software SPSS, se presentan clasificados en dos grupos denominados: Usos de Internet para actividades académicas y uso de Internet para actividades de entretenimiento.

El grupo correspondiente a Usos de Internet para actividades académicas, contiene a los estudiantes que utilizan Internet con fines de aprendizaje.

El grupo correspondiente a Usos de Internet para actividades de entretenimiento, corresponde a aquellos estudiantes que más dan uso a Internet con fines sociales y de entretenimiento.

3.2.3.5 Análisis Discriminante.

Es una técnica estadística que nos permite conocer las variables que nos permiten diferenciar (discriminar) a los grupos y cuántas de estas variables son necesarias para alcanzar la mejor clasificación posible (Ruiz & Pardo, 2001).

Esta técnica se utiliza para predecir una variable de respuesta categórica. A diferencia de los modelos lineales generalizados, supone que la variable independiente sigue una distribución normal multivariante (Han, Kamber, & Pei, 2012). Puede utilizarse como complemento al análisis clúster, para ello la variable dependiente es la que resulta del análisis clúster (Torres, 2012).

Para identificar la clasificación de los estudiantes en el *uso de Internet para actividades académicas*, la asociación en dos clúster tiene un nivel de exactitud del 98.4%, mientras que la clasificación de los estudiantes en el uso de Internet para actividades de

entretenimiento la clasificación en tres clúster muestra una exactitud del 97.1% (Ver anexos 7 y 12).

3.2.3.6 Análisis de Varianza.

Esta técnica permite determinar la igualdad de las varianzas de dos poblaciones; además compara si las medias de más de dos poblaciones son iguales entre sí.

Se lo conoce como ANOVA y tiene tres supuestos:

1. Las poblaciones a las que pertenecen las muestras tienen distribución normal.
2. Las poblaciones tienen desviaciones estándar iguales.
3. Las poblaciones son independientes.

Para verificar el supuesto de homogeneidad de la varianza se utiliza el estadístico de Levene, que en caso de tener un valor significativo ($p < 0,05$) e indica que las varianzas son significativamente diferentes (Torres, 2012).

3.2.3.7 Regresión Logística.

Regresión Logística es un tipo de análisis que ayuda a predecir el resultado de una variable dependiente dicotómica que generalmente puede tener dos valores (0 y 1), y las variables predictoras o independientes que pueden ser cuantitativas (categóricas o continuas). Se utiliza para observar la probabilidad de un suceso (Torres, 2012).

El propósito de la Regresión Logística es:

- Predecir la probabilidad de que ocurra un evento.
- Determinar qué variables intervienen para aumentar o disminuir la probabilidad de un evento.

La Regresión Logística Binaria es uno de los modelos más utilizados y de mayor interés para analizar probabilidades.

En nuestro estudio, por medio de la Regresión Logística podemos comprobar las hipótesis planteadas, utilizando las siguientes pruebas:

3.2.4 Procedimiento para Evaluación y presentación de datos.

En esta fase damos respuesta a las preguntas de investigación planteadas, se evalúan los resultados obtenidos en el análisis clúster con el fin de determinar el mejor modelo que se ajuste a nuestros datos.

Por ello utilizamos la técnica de regresión logística binaria y multinomial según es el caso que dé respuesta a las hipótesis planteadas y utilizando las variables dependientes e independientes respectivas.

La Regresión Logística Binaria se aplicó para las hipótesis 1, 3 y 5, mientras que la Regresión Logística Multinomial para la hipótesis 4.

Una vez aplicada la técnica respectiva, se procede a comprobar o rechazar las hipótesis utilizando las medidas de bondad de ajuste: Omnibus, Hosmer y Lemeshow, estadístico R cuadrado de Nagelkerke, estadístico de Wald y la prueba de χ^2 .

CAPÍTULO IV: MUESTRA DE RESULTADOS

4.1 Generalidades sobre la muestra.

La ESPOL, en su oferta académica cuenta con las siguientes carreras:

Tabla 4: Oferta Académica de la ESPOL

Facultades	Carreras
Ciencias de la Tierra	Postgrados Maestría en Arqueología del Neotrópico Pregrado Ingeniería Civil Ingeniería de Minas Ingeniería de Petróleo Ingeniería en Geología
Electricidad y Computación	Postgrados Maestría en Sistemas Eléctricos de Potencia Maestría en Automatización y Control Industrial Maestría en Sistemas de Información Gerencial Maestría en Seguridad Informática Aplicada Maestría en Telecomunicaciones Pregrado ING. ELECTRÓNICA Y TELECOMUNICACIONES ING. EN ELECTRIC. ELECTRÓNICA Y AUTOMATIZACIÓN INDUSTRIAL ING. EN ELECTRICIDAD. POTENCIA INGENIERÍA EN CIENCIAS COMPUTACIONALES INGENIERÍA EN TELEMÁTICA LICENCIATURA EN REDES Y SISTEMAS OPERATIVOS LICENCIATURA EN SISTEMA DE INFORMACIÓN (FIEC)
Marítima, Ciencias Biológicas, Oceánicas y Recursos Naturales	Postgrados Maestría en Cambio Climático Pregrado Biología Marina Ingeniería en Acuicultura Ingeniería Naval Ingeniería Oceánica y Ciencias Ambientales Licenciatura en Turismo
Mecánica y Ciencias de la Producción	Pregrado Ingeniería Agrícola y Biológica Ingeniería en Alimentos Ingeniería Mecánica Ingeniería y Administración de la Producción Industrial Tecnología en Agricultura
Ciencias Sociales y Humanísticas.	Postgrados Maestría en Economía y Administración de Empresas

	Maestría en Gestión del Talento Humano Maestría en Finanzas Pregrado Economía Con Mención en Gestión Empresarial Ingeniería Comercial y Empresarial Ingeniería en Marketing, Comunicación y Ventas Ingeniería en Negocios Internacionales
Ciencias Naturales y Matemáticas.	Departamento de Matemáticas Postgrados Maestría en Seguros y Riesgos Financieros Maestría en Educación con Mención en Enseñanza de la Matemáticas Maestría en Gestión de la Productividad y la Calidad Maestría en Control de Operaciones y Gestión Logística Pregrado INGENIERÍA EN AUDITORÍA Y CONTADURÍA PÚBLICA AUTORIZADA INGENIERÍA EN ESTADÍSTICA INFORMÁTICA INGENIERÍA EN LOGÍSTICA Y TRANSPORTE Departamento de Física Postgrados Maestría en Enseñanza de la Física Departamento de Química y Ciencias Ambientales Postgrados Maestría en Ciencias Ambientales Maestría en Manejo Integral de Laboratorios de Desarrollo Pregrado Ingeniería Química

La **Escuela de Diseño y Comunicación Visual**, ofrece carreras de pregrado en el área de computación y diseño gráfico.

Escuela	Carreras
Diseño y Comunicación Visual	Postgrados
	Maestría En Diseño y Gestión de Marca
	Pregrado
	Licenciatura en Comunicación Social
	Licenciatura en Diseño Gráfico y Publicitario
	Licenciatura en Diseño Web y Aplicaciones Multimedia
	Licenciatura en Diseño y Producción Audiovisual

Fuente: <http://www.espol.edu.ec/espol/main.jsp?id=8>

Elaborado por: ESPOL

Referente a los datos relevantes sobre la muestra de la población, se ha obtenido lo siguiente:

4.1.1 Datos relevantes sobre la población: Género, edades, ingresos familiares.

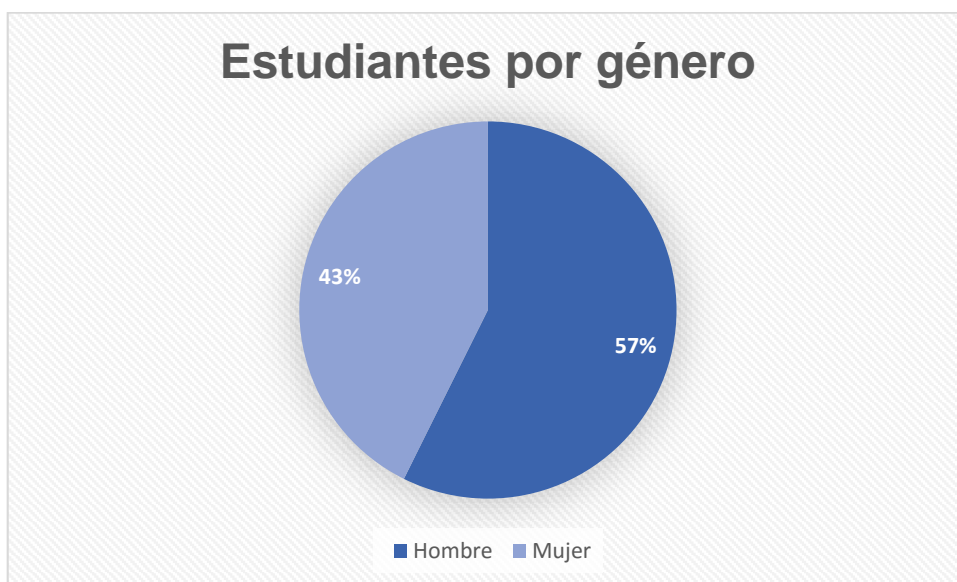


Figura 7: Porcentaje de estudiantes por género

Fuente: Elaboración del autor

La figura 7, muestra la evidencia que la proporción por género es mayor en los hombres a la de mujeres.

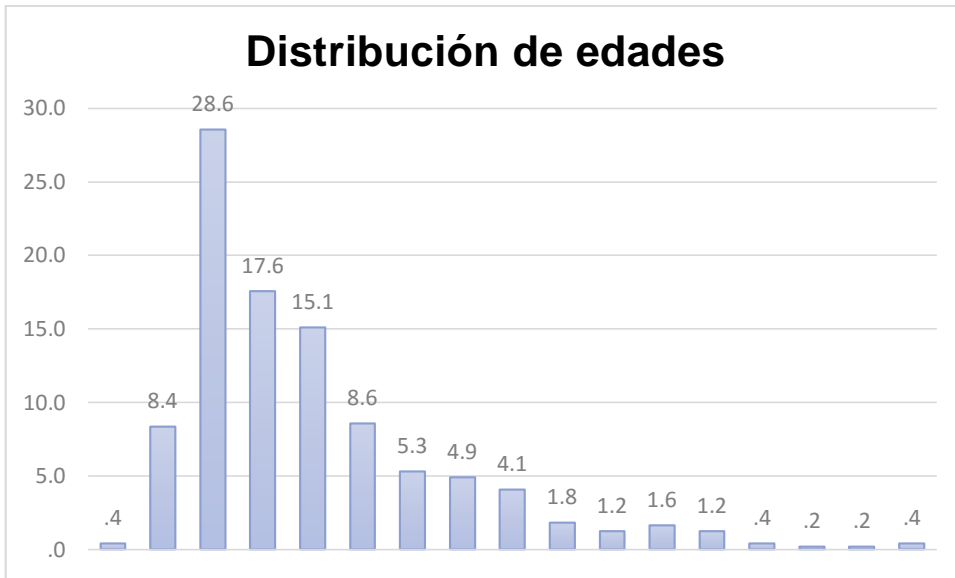


Figura 8: Distribución de edades
Fuente: Elaboración del autor

Se encontró que la edad promedio de los estudiantes es 19 años, figura 8. Esta población se caracteriza por formar parte del grupo llamado nativos digitales.

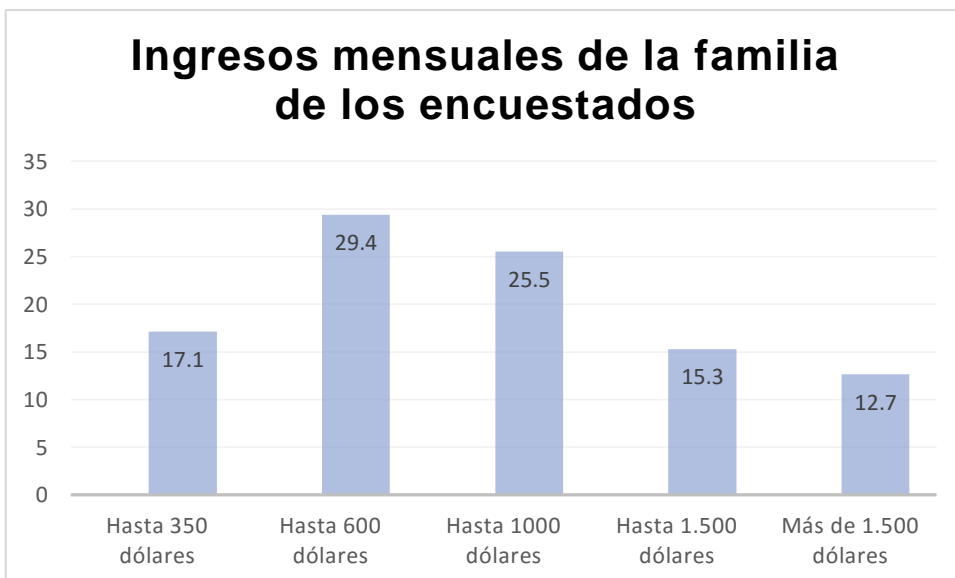


Figura 9: Ingresos mensuales de los familiares de los encuestados
Fuente: Elaboración del autor

La figura 9 muestra que el mayor porcentaje 54,9% de los estudiantes correspondientes al nivel de ingresos mensuales de sus familias se concentra en quienes perciben un rango entre 600 a 1000 dólares.

4.1.2 Datos relevantes sobre el conocimiento y el uso de Internet.

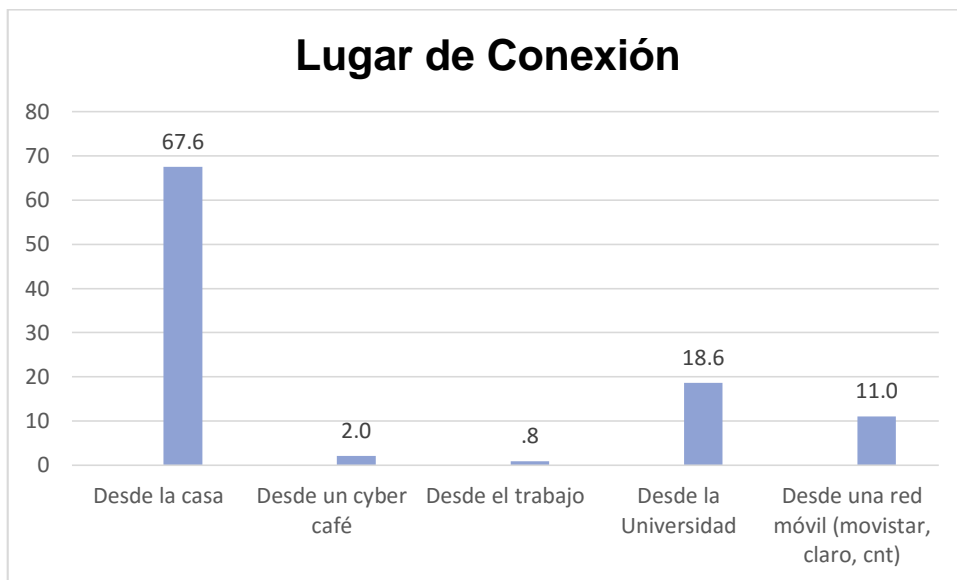


Figura 10: Lugar de Conexión

Fuente: Elaboración del autor

Lo que refiere al lugar de conexión preferida por el estudiante, el 67,6% tiene predilección por acceder desde su casa, figura 10.

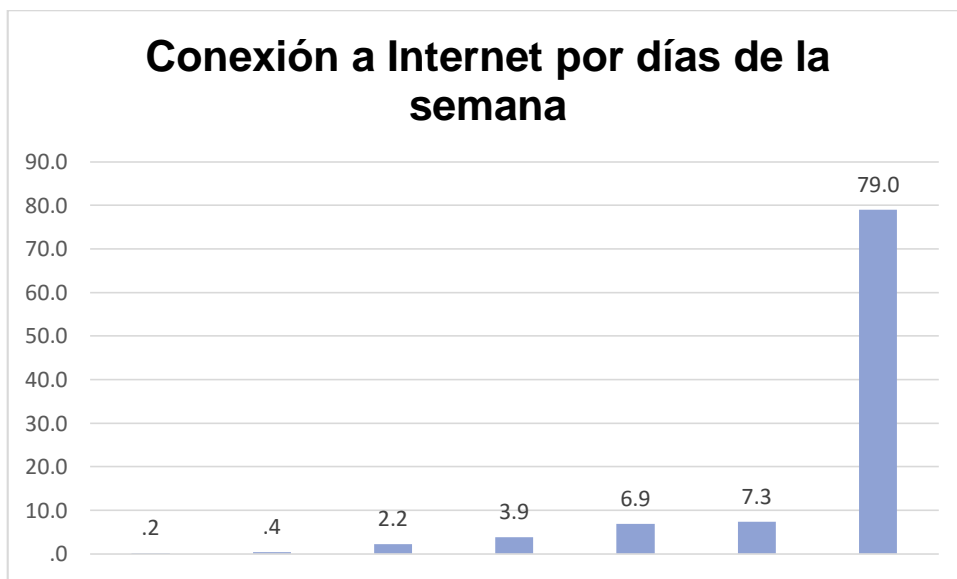


Ilustración 11: Días a la semana que se conecta a Internet

Fuente: Elaboración del autor

Podemos observar que, el 79% de los estudiantes se conectan los 7 días de la semana; en contraste al 0,2% que suele hacerlo 1 vez a la semana. Lo que permite inferir que la mayoría de estudiantes pasa conectado a Internet y a simple vista podemos apreciar que mientras mejores ingresos económicos tenga la familia del estudiante, su acceso a Internet se incrementa.

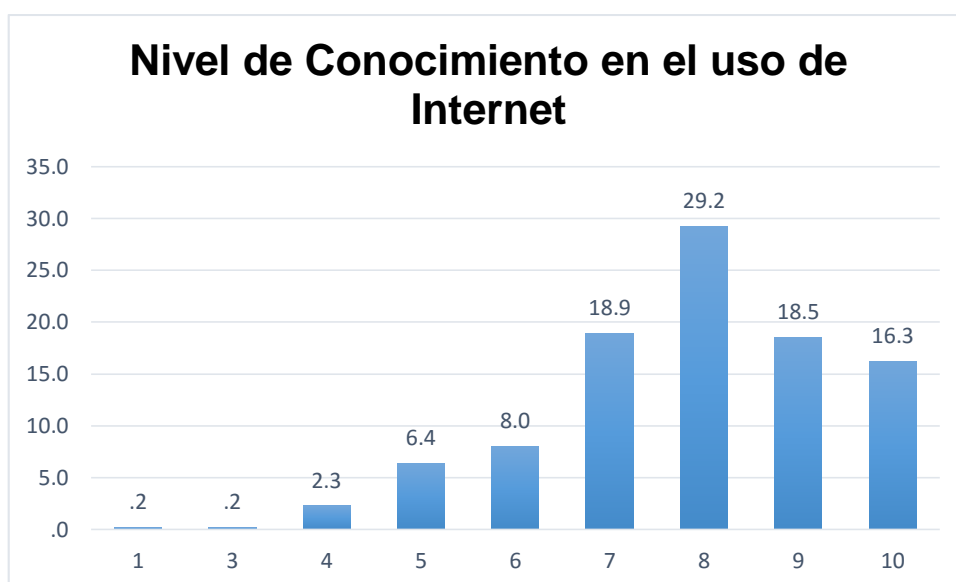


Figura 12: Nivel de conocimiento en el uso de Internet

Fuente: Elaboración del autor

En la figura 12, se observa que los niveles de conocimiento en el uso de Internet (1 a 10), el 82,9% tiene conocimientos elevados, es decir su nivel de conocimiento se encuentra entre 7 y 10.

Tabla 5: Horas de conexión de los estudiantes

Horas de conexión	Porcentaje
1,00	1,0%
2,00	9,4%
3,00	13,1%
4,00	12,2%
5,00	14,7%
6,00	7,1%
7,00	5,5%
8,00	6,5%
9,00	4,3%
10,00	9,4%

12,00	5,7%
14,00	2,7%
15,00	8,4%
Total	100,0

Fuente: Elaboración del autor

La tabla 5, nos muestra que un 63% de los estudiantes se conecta entre 1 y 7 horas, mientras que 37% lo hace entre 8 y 15 horas diarias.

En la tabla 6, podemos observar la distribución de estudiantes por años de experiencia en uso de Internet, y nos refleja que un 80,7 % tiene entre 2 y 9 años de experiencia, y un 19,3 % entre 10 y 16 años de uso de Internet. Por lo que el promedio en años de experiencia de usos de Internet de los estudiantes de la ESPOL es de 5 años con un 16,7%

Tabla 6: Años de experiencia utilizando Internet

Años de Experiencia	Porcentaje
2	1,8%
3	6,7%
4	14,9%
5	16,7%
6	14,9%
7	11,4%
8	9,2%
9	9,8%
10	9,0%
11	2,0%
12	1,4%
13	1,0%
14	0,6%
15	0,2%
16	0,2%
Total	100,0

Fuente: Elaboración del autor

Tabla 7: Accesos semanales a la plataforma virtual de la Universidad

Número de accesos por semana	Porcentaje
0	1,0
1	3,3
2	8,0
3	10,8
4	9,4
5	17,1
6	18,8
7	18,8
8	1,2
9	,8
10	8,2
12	1,0
13	,2
14	1,4
Total	100,0

Fuente: Elaboración del autor

Se puede observar que un 37,6% de los estudiantes acceden entre 6 y 7 veces semanales, seguido por un 17,1% que lo realiza 5 veces. Un 10,8% accede 3 veces a la semana y los demás días en un menor porcentaje, ver tabla 7.

4.2 Relaciones entre variables.

- Existe relación entre la edad y el nivel de conocimiento del manejo de Internet (R de Pearson = 0,004; Coeficiente de Determinación = 1,69%; $p = 0,004 < 0,05$), se rechaza la hipótesis nula concluyendo que la edad sí tiene que ver con el conocimiento del manejo de Internet.
- Existe relación entre la edad y el tiempo que se conecta a Internet (R de Pearson = 0,311; Coeficiente de Determinación = 9,67%; $p = 0,000 < 5$), se rechaza la hipótesis nula concluyendo que la edad sí influye en el tiempo que se conecta a Internet.
- Existe relación entre la edad y el número de horas de uso de la biblioteca virtual de la Universidad (R de Pearson = -0,107; Coeficiente de Determinación = 1,14%; $p = 0,018 < 0,05$), se rechaza la hipótesis nula y se concluye que la edad sí está relacionada con el uso de la biblioteca virtual de la Universidad.

- Existe relación entre la edad y el número de horas a la semana que chatea por diversión el estudiante (R de Pearson = - 0,134; Coeficiente de Determinación = 1,79%; $p = 0,003 < 0,05$), se rechaza la hipótesis nula y se concluye que la edad sí influye en las horas de chat por semana de los estudiantes.
- Existe relación entre la edad y el número de horas a la semana que accede a redes sociales (R de Pearson = - 0,077; Coeficiente de Determinación = 0,59%; $p = 0,087 < 0,05$), por lo que se rechaza la hipótesis nula y se concluye que los estudiantes más jóvenes son los que usan más las redes sociales.
- Se encuentra relación entre la edad y el uso de juegos en línea por hora a la semana (R de Pearson = - 0,103; Coeficiente de Determinación = 1,06%; $p = 0,023 < 0,05$), por lo tanto se rechaza la hipótesis nula y se concluye la edad está relacionada con las horas que usan para juegos en línea.
- Existe relación entre la edad y las horas que utiliza para descargar música, videos y programas (R de Pearson = - 0.102; Coeficiente de Determinación = 1,04%; $p = 0,024 < 0,05$) por tanto se rechaza la hipótesis nula y se concluye que la edad sí está relaciona con el número de horas que utiliza para descargar música, videos y programas.

4.3 Perfil uso de Internet para actividades académicas.

Para determinar el Perfil de uso de Internet y la relación con las actividades académicas de los estudiantes se procedió a realizar 3 pasos:

1. Análisis factorial para reducir variables.
2. Clasificación de los estudiantes a través del procedimiento análisis de Conglomerados de K-medias.
3. Verificación de la clasificación a través del análisis de Varianza y Discriminante.

Análisis factorial para reducir variables

Para identificar el mayor uso que tienen los estudiantes de Internet en sus actividades académicas se procedió a identificar las variables significativas, es decir aquellas cuyos valores son más representativos.

Para ello se utilizó análisis factorial, que es una de las técnicas de reducción de datos que sirve para encontrar grupos homogéneos de variables a partir de un conjunto numeroso de ellas.

Tabla 8: Uso de Internet en Actividades Académicas

	Estadísticos									
	A_Ingre	A_Cons	A_Consumo	A_Recurso	A_Videos	A_Foros	A_Post	A_Horas	A_Horas	A_Biblioteca
	soPlataforma	ultasProfesor	ltasEstudiantes	osEducativos	sAcademicos	osVirtuales	sAcademicos	ChatAcademico	asBusqueda	otecaVirtual
Media	6.00	2.00	10.00	7.00	7.00	1.00	2.00	10.00	20.00	3.00

Fuente: Elaboración del Autor

Gráficamente podemos observar:

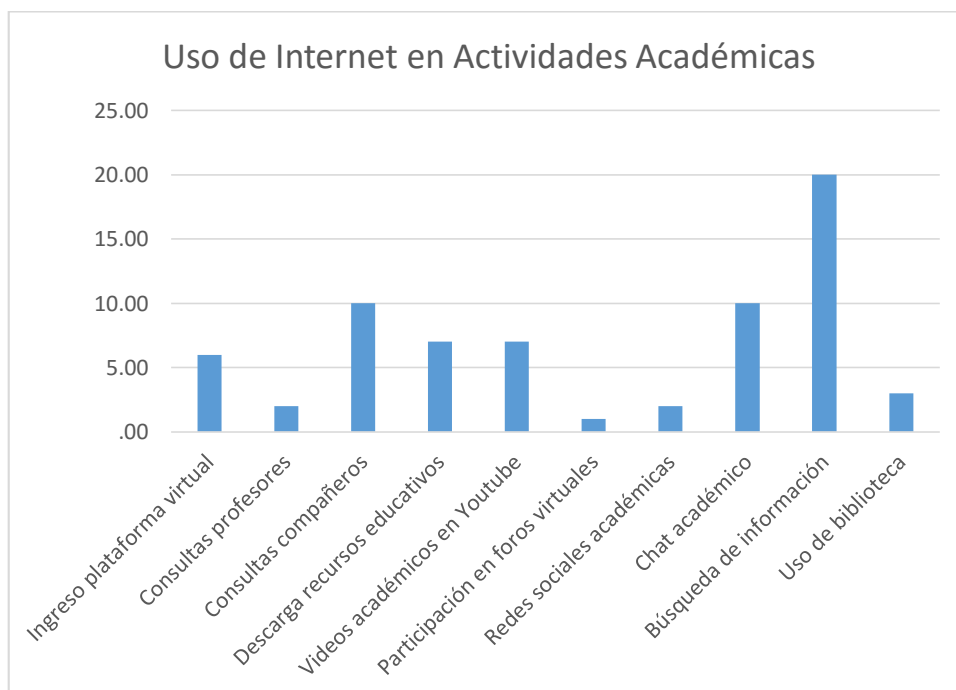


Figura 13: Uso de Internet en Actividades Académicas

Fuente: Elaboración del Autor

Las variables que destacan corresponden a los siguientes elementos:

Búsqueda de información

Chat académico

Consultas compañeros

Descarga de recursos educativos

Videos académicos en Youtube

Se clasifican los estudiantes y por medio del análisis de *Conglomerados de k – medias*, se obtienen clasificaciones de dos, tres, cuatro y cinco grupos para escoger el más adecuado (Anexos 4, 5, 6 y 7). El análisis de conglomerados de k – medias, como técnica multivariante permite agrupar las variables en función de su parecido o similitud presentada.

Procedemos a realizar la comprobación a través del análisis discriminante que ayuda a identificar las características que discriminan a dos o más grupos y permite crear una función que distingue con mayor precisión los porcentajes asignados a uno u otro grupo (Anexo 5).

Tabla 9: Clasificación del Discriminante grupos de Actividades Académicas

Grupos de Actividades Académicas	Porcentaje
Dos	98.4%
Tres	97.5%
Cuatro	97.5%
Cinco	97.8%

Fuente: Elaboración del Autor

La clasificación mejor realizada corresponde a la clasificación en dos grupos: Tradicional y Tecnológico, la (Figura 14) gráficamente nos presenta la descripción de los mismos.

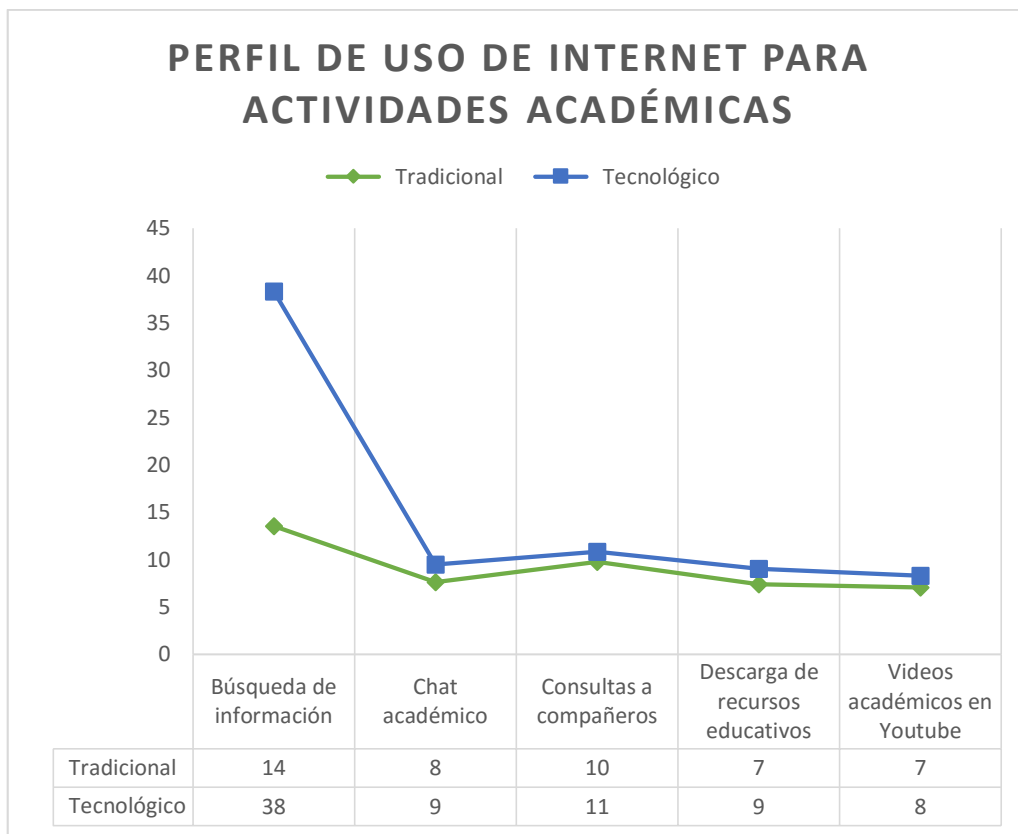


Figura 14: Perfil de uso de Internet para Actividades Académicas

Fuente: Elaboración del Autor

El *perfil tradicional*, corresponde al 26,35% de la clasificación, los valores son menores al conglomerado tecnológico. Este grupo se caracteriza por que sus valores son menores en la búsqueda de información, chat con fines académicos, consultas a compañeros, descarga de recursos educativos y uso de Youtube con fines académicos. Este grupo busca más información. El 59,7% son hombres y el 40,3% mujeres.

El *perfil tecnológico*, corresponde al 73,67% de la población, los valores son mayores al grupo tradicional. Este grupo se caracteriza por buscar más información, chatear con fines académicos, realizar consultas a sus compañeros, descargar recursos educativos y ver videos académicos en Youtube. El 56,5% son hombres y el 43,5% son mujeres.

4.4 Perfil uso de Internet para actividades de entretenimiento.

Para proceder a este análisis, tomamos en cuenta las variables que corresponden a las actividades de entretenimiento de nuestro proyecto, y como en el proceso anterior, procedemos a tomar aquellos valores más significativos:

Tabla 10: Uso de Internet en Actividades de Entretenimiento

Estadísticos					
	E_HorasChatPersonal	E_Redessociales	E_JuegosLinea	E_DescargaMusica	E_Videos
Mediana	10.00	10.00	.00	4.00	8.00

Fuente: Elaboración del Autor

Gráficamente se observa:

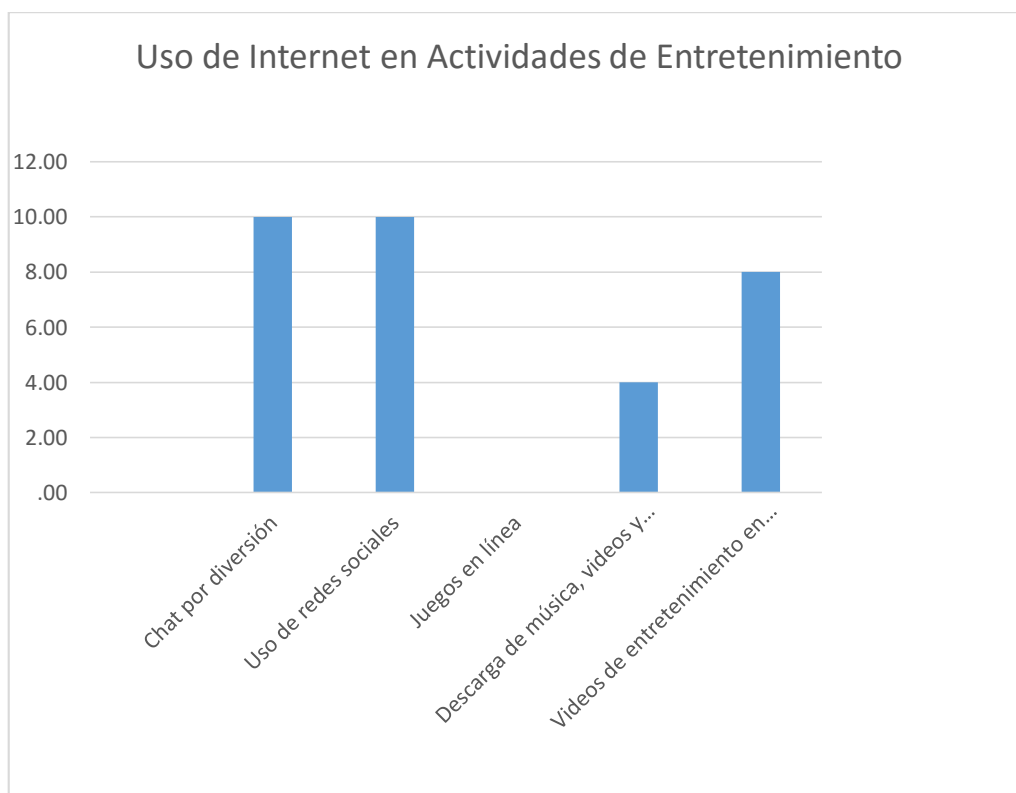


Figura 15: Uso de Internet en Actividades de Entretenimiento

Fuente: Elaboración del Autor

Finalmente, las variables que destacan corresponden a:

Chat por diversión

Uso de redes sociales

Videos de entretenimiento en Youtube

Se clasifican los estudiantes y por medio del análisis de *Conglomerados de k – medias*, se obtienen clasificaciones de dos, tres, cuatro y cinco grupos para escoger el más adecuado (Anexos 6, 7, 8 y 9).

Procedemos a realizar la comprobación a través del análisis discriminante (Anexo 10) y nos arroja la siguiente tabla:

Tabla 11: Clasificación del Discriminante grupos de Actividades de Entrenamiento

Grupos de Actividades de Entretenimiento	Porcentaje
Dos	96.6%
Tres	97.1%
Cuatro	96.9%
Cinco	96.9%

Fuente: Elaboración del Autor

La clasificación mejor realizada corresponde a la clasificación en tres grupos: Alto, medio y bajo (Figura 16).

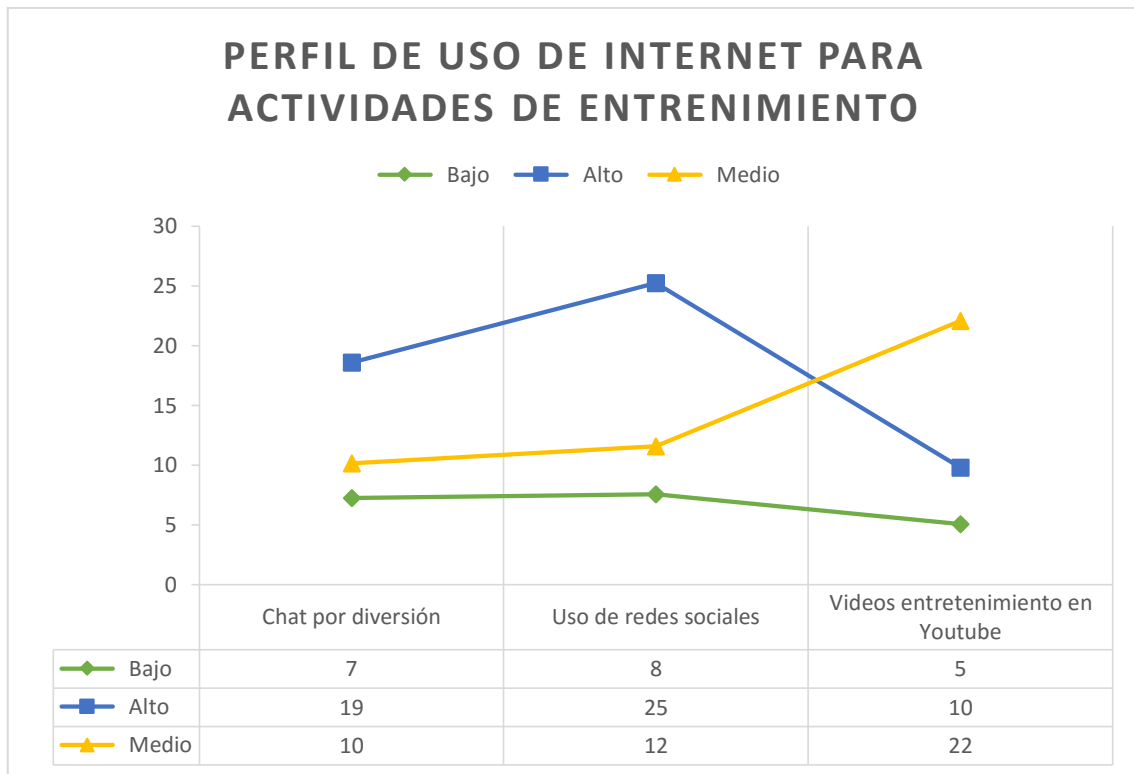


Figura 16: Perfil de uso de Internet para Actividades de Entretenimiento
Fuente: Elaboración del Autor

El *perfil Alto*, corresponde al 62,44%, se caracteriza por la alta utilización de redes sociales y chat por diversión (19 horas a la semana), aunque mira menos videos para entretenimiento en Youtube. 53,2% corresponde a hombres y 43,8% mujeres.

El *perfil Medio*, corresponde a un 20,40% de la población, se destaca en un alto valor de observación de videos para entretenimiento en Youtube, mediano uso de redes sociales y bajo uso del chat por diversión, apenas 10 horas semanales. 56% de hombres y 44% de mujeres.

El *perfil Bajo*, corresponde a un 17,16%, semanalmente chatean 7 horas, la utilización de redes sociales es menor y usan menos Youtube para visualizar videos para entretenimiento. 63,1% hombres y 36,9% mujeres.

4.5 Fase de Evaluación.

En esta fase del proceso de minería de datos, la información obtenida es analizada y evaluada por medio de Regresión Logística que permita comprobar las hipótesis planteadas para este proyecto.

4.5.1 Evaluación del perfil uso de Internet y actividades académicas.

El perfil de uso de Internet y actividades académicas corresponde a la clasificación de los estudiantes en base a la definición de perfiles del uso de Internet para actividades académicas (clúster en 2 categorías).

Se desea comprobar la hipótesis 1: “El nivel de ingresos determina como se utiliza Internet para el aprendizaje”.

El modelo de comprobación de hipótesis aplica la Regresión Logística Binomial, se utiliza como variable dependiente el grupo definido de usos de Internet en Actividades Académicas que cuenta con dos niveles. Para esto la variable dependiente se recategorizó como una variable Dummy o binaria, y se toma como valor de referencia los ingresos mensuales de hasta 350 dólares (Ver anexo 14).

Para realizar la comprobación del presente modelo se utilizó las pruebas de Omnibus, Hosmer y Lemeshow.

La prueba Omnibus, nos ayuda a verificar que al menos uno de los coeficientes del modelo sea diferente de 0, los valores χ^2 presentan un valor de ($\chi^2= 0.714$, $p = 0.950$) por lo que no es un modelo significativo ($p > 0.05$), lo que manifiesta que el modelo no es aceptable.

El estadístico R cuadrado de Nagelkerke tiene un valor de 0.002 (0.2%), que no es cercano a 0 y su valor corresponde al porcentaje de varianza explicada por el modelo.

La prueba de bondad de ajuste de Hosmer y Lemeshow alcanza un valor de ($\chi^2 = 0$, $p = 1$), lo que verifica la hipótesis 1: Los valores observados son iguales a los valores predichos. Por lo que no se rechaza H_0 .

Para verificar los coeficientes de la variable independiente se utiliza la prueba de Wald (Tabla 9), sin embargo, las variables no muestran valores significativos ($p > 0.05$).

Tabla 12: Coeficientes del modelo de regresión logística para el perfil de uso de Internet y actividades académicas

		Variables en la ecuación							
		B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95% para EXP(B)	
								Inferior	Superior
Paso 1a	Ingresos			.708	4	.950			
	Ingresos(1)	-.137	.318	.186	1	.666	.872	.467	1.627
	Ingresos(2)	-.219	.325	.454	1	.500	.804	.425	1.518
	Ingresos(3)	-.219	.363	.363	1	.547	.804	.394	1.637
	Ingresos(4)	-.021	.392	.003	1	.957	.979	.454	2.111
	Constante	1.163	.256	20.616	1	.000	3.200		

a. Variables introducidas en el paso 1: Ingresos

Fuente: Elaboración del autor

El modelo de regresión logística se define:

$$\Pr(Ac_Aca) = \frac{1}{1 + e^{-(1.1163 - (-0.137(\text{ingresos}(1)) - (-0.219(\text{ingresos}(2)) - (-0.219(\text{ingresos}(3)) - (-0.021(\text{ingresos}(4))))))}}$$

Podemos observar que se muestra una diferencia estadística no significativa, por lo que las constantes no explican si hay pertenencia a la variable dependiente, porque lo que el nivel de ingresos no determina si los estudiantes utilizan Internet para el aprendizaje.

Finalmente, se aplica la prueba χ^2 , para medir la diferencia entre los valores esperados y los valores observados, los resultados presentan un valor no significativo ($\chi^2 = 0.709$, $p = 0.950$; $p > 0.05$) con 4 grados de libertad y así se verifica que no existe incidencia significativa entre las variables dependientes e independientes, por lo que se rechaza H_1 (Anexo 14).

4.5.2 Evaluación del perfil de uso de Internet para actividades de entretenimiento.

Este perfil se elaboró considerando la clasificación de 3 grupos de los estudiantes y el uso de Internet para actividades de entretenimiento. Este resultado se clasifica en: Alto, Medio y Bajo.

En esta sección se da respuesta a la Hipótesis 2: “El nivel de ingresos determina como se utiliza Internet para entretenimiento”.

La bondad de ajuste del modelo se verificó por medio del valor χ^2 del logaritmo de la verosimilitud ($\chi^2= 11.088$ $p = 0.197$), los valores de Pearson ($\chi^2= 0$) y la Desviación ($\chi^2= 0$), que son valores significativos ($p > 0.05$). La varianza que explica el modelo a través del valor Nagelkerke = 0.027 (2.70 %) (Anexo 15). La prueba de Wald, en la tabla 10 nos muestra que en el modelo Bajo, la variable de ingresos 1, ($p = 0.017$; $p > 0.05$), en el resto de niveles se produce lo contrario.

En el modelo el OR de pertenecer al perfil de entretenimiento Medio con respecto al perfil Alto es 0.919 veces mayor al estudiante que pertenece al nivel 3 respecto del nivel 5 (OR = 0.919, (IC 95% 0.444 – 1.901), $p = 0.819$), es 0.787 veces mayor cuando el estudiante pertenece al nivel de ingresos 2 respecto del nivel 5 (OR = 0.787, (IC 95% 0.387 – 1.602), $p = 0.509$), es 0.631 veces mayor cuando el estudiante pertenece al nivel de ingresos 4 respecto del nivel 5 (OR = 0.631, (IC 95% 0.265 – 1.501), $p = 0.297$), es 0.439 veces mayor cuando el estudiante pertenece al nivel de ingresos 1 respecto del nivel 5 (OR = 0.439, (IC 95% 0.183 – 1.048).

Tabla 13: Coeficientes del modelo de regresión logística multinomial para el perfil de Uso de Internet para actividades de entretenimiento

Estimaciones de los parámetros									
Número inicial de casosa		B	Error típ.	Wald	gl	Sig.	Exp(B)	para Exp(B)	
								Límite inferior	Límite superior
Medio	Intersección	-.838	.299	7.850	1	.005			
	[Ingresos=1]	-.824	.445	3.437	1	.064	.439	.183	1.048
	[Ingresos=2]	-.239	.363	.435	1	.509	.787	.387	1.602
	[Ingresos=3]	-.085	.371	.052	1	.819	.919	.444	1.901
	[Ingresos=4]	-.461	.442	1.086	1	.297	.631	.265	1.501
	[Ingresos=5]	0	.	.	0
Bajo	Intersección	-1.414	.372	14.468	1	.000			
	[Ingresos=1]	.061	.471	.017	1	.896	1.063	.422	2.677
	[Ingresos=2]	-.239	.452	.280	1	.597	.787	.325	1.909
	[Ingresos=3]	.259	.442	.343	1	.558	1.295	.545	3.080
	[Ingresos=4]	.574	.462	1.543	1	.214	1.775	.718	4.391
	[Ingresos=5]	0	.	.	0

a. La categoría de referencia es: Alto.

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

Fuente: Elaboración del autor

En el segundo modelo el OR, de pertenecer al perfil de entretenimiento Bajo, con respecto al Perfil de entretenimiento Alto, se observa que para todos los niveles de ingresos no existe una relación significativa ($p > 0.05$).

4.5.3 Evaluación de la incidencia del uso de Internet para actividades académicas sobre el rendimiento académico.

Realizaremos la evaluación de la hipótesis 3: “El uso de la tecnología en el aprendizaje incide en el rendimiento académico”.

Para construir el modelo se creó la variable “Rendimiento_Academico”, esta variable resulta de la diferencia de las variables SA_AsignaturasMatriculadas y SA_AsignaturasAprobadas, lo que arrojó un resultado de una variable categórica de dos niveles:

1 = “Aprobado”

0 = “Reprobado”

La variable “Rendimiento_Academico” es la dependiente y la variable “Cluster_Academico_2” es la independiente.

Se aplica Regresión Logística Binaria porque la variable dependiente tiene dos niveles.

La prueba de bondad de ajuste Omnibus nos da como resultado ($\chi^2 = 0.584$, $p = 0.445$, $p > 0.05$), por lo que se define que el modelo no tiene capacidad predictora (Anexo 16).

El estadístico R cuadrado de Nagelkerke tiene un valor de 0.002 (0.2%), que no es cercano a 0 y su valor corresponde al porcentaje de varianza explicada por el modelo.

La prueba de bondad de ajuste de Hosmer y Lemeshow alcanza un valor de ($\chi^2 = 0$, $p = 1$), lo que verifica H_0 : Los valores observados son iguales a los valores predichos. Por lo que no se rechaza H_0 .

Para verificar los coeficientes de la variable independiente se utiliza la prueba de Wald, sin embargo, las variables no muestran valores significativos ($p > 0.05$).

Tabla 14: Coeficientes de la regresión logística binaria entre el Uso de Internet en actividades académicas sobre el rendimiento académico

		Variables en la ecuación							
		B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95% para EXP(B)	
								Inferior	Superior
Paso 1a	Cluster_Academico_2(1)	-.160	.210	.581	1	.446	.852	.564	1.287
	Constante	.490	.181	7.302	1	.007	1.633		

a. Variable(s) introducida(s) en el paso 1: Cluster_Academico_2.

Fuente: Elaboración del autor

El OR no presenta relaciones significativas, como resultado se rechaza la hipótesis 3 y se deduce que no existe incidencia en el uso de Internet en actividades académicas sobre el rendimiento académico.

Finalmente, se aplica la prueba χ^2 , para medir la diferencia entre los valores esperados y los valores observados, los resultados presentan un valor no significativo ($\chi^2 = 0.709$, $p = 0.950$; $p > 0.05$) y así se verifica que no existe incidencia significativa entre las variables dependientes e independientes, por lo que se rechaza H_1 . (Anexo 16).

4.5.4 Evaluación de la incidencia del uso de Internet para actividades de entretenimiento sobre el rendimiento académico.

Para comprobar la hipótesis 4: “El uso de la tecnología para entretenimiento incide en el rendimiento académico”.

Se construye un modelo de Regresión Logística Binomial, la variable Dependiente la constituye “Rendimiento_Academico”, formada por los elementos 1 = “Aprobado” y 0 = “Reprobado”; la variable Independiente es “Cluster_Entretenimiento_3”, que incluye: Alto, Medio y Bajo referente a los usos de Internet para actividades de entretenimiento en los estudiantes de la ESPOL.

La prueba de bondad de ajuste Omnibus nos da como resultado ($\chi^2 = 0.165$, $p = 0.685$, $p > 0.05$), por lo que se define que el modelo no tiene capacidad predictora (Anexo 17).

El estadístico R cuadrado de Nagelkerke tiene un valor de 0.000, y su valor corresponde al porcentaje de varianza explicada por el modelo.

La prueba de bondad de ajuste de Hosmer y Lemeshow alcanza un valor de ($\chi^2 = 0.170$, $p = 0.681$), con un grado de libertad lo que verifica H_0 : Los valores observados son iguales a los valores predichos. Por lo que no se rechaza H_0 .

Para verificar los coeficientes de la variable independiente se utiliza la prueba de Wald, sin embargo, las variables no muestran valores significativos ($p > 0.05$).

Tabla 15: Coeficientes de la regresión logística binaria entre el Uso de Internet en actividades de entretenimiento sobre el rendimiento académico

		Variables en la ecuación						I.C. 95% para EXP(B)	
		B	E.T.	Wald	gl	Sig.	Exp(B)	Inferior	Superior
Paso 1a	Cluster_Entrenamiento_3	.049	.120	.164	1	.685	1.050	.830	1.328
	Constante	.296	.207	2.059	1	.151	1.345		

Fuente: Elaboración del autor

El OR de este modelo es menor y por lo tanto no significativo.

Se aplica la prueba χ^2 , para medir la diferencia entre los valores esperados y los valores observados, los resultados presentan un valor no significativo ($\chi^2 = 0.332$, $p = 0.847$; $p > 0.05$) y así se verifica que no existe incidencia significativa entre las variables dependientes e independientes, por lo que se rechaza H_0 y podemos definir que no existe incidencia entre el uso de Internet en actividades de entretenimiento sobre el rendimiento académico (Anexo 16).

CAPÍTULO V: DISCUSIÓN DE RESULTADOS

Este apartado corresponde a la presentación de Conocimiento de las fases de KDD para Minería de Datos.

Internet ha experimentado un rápido crecimiento en los lugares que hay acceso y la brecha digital prácticamente no existe. Sin embargo, esta penetración implica que las competencias digitales de los estudiantes no se enfocan del todo hacia el correcto uso de la WEB.

En el presente proyecto podemos concluir que no necesariamente el nivel de ingresos de las familias de los jóvenes de la ESPOL, influye en el uso de Internet tanto para actividades académicas, como actividades de entretenimiento. Un tema interesante de reflexión es la generación actual de jóvenes, conocidos como generación Z, generación caracterizada por vivir conectados a la WEB, pero no explotar su capacidad. Para (Castaño M, 2011) la introducción de Internet en el aprendizaje es eficaz solamente para mejorar el rendimiento cuando se explota su capacidad de cambio.

En nuestra muestra la edad promedio es 21 años y 140 jóvenes tienen 19 años de edad, su grupo etario corresponde a la Generación Z⁶. De este grupo de jóvenes, podemos evidenciar que 331 jóvenes se conectan a Internet desde su casa, igualmente estos jóvenes tienen acceso a la WEB los 7 días de la semana, un 15% se conecta o pasa conectado en promedio de 5 horas diarias; es decir, prácticamente pasan conectados. Este grupo, el 16.7%, tiene una experiencia de uso de Internet de unos 5 años.

5.1 Discusión de resultados.

Respecto a la **Hipótesis 1**: Se observó que existe diferencia estadística no significativa, por lo que las constantes no explican si hay pertenencia a la variable dependiente, porque lo que el nivel de ingresos no determina si los estudiantes utilizan internet para el aprendizaje.

⁶ Nacidos desde 1995

Al aplicar la prueba χ^2 , para medir la diferencia entre los valores esperados y los valores observados, los resultados presentan un valor no significativo ($\chi^2 = 0.709$, $p = 0.950$; $p > 0.05$) con 4 grados de libertad y así se verifica que no existe incidencia significativa entre las variables dependientes e independientes, por lo que se rechaza H1 (Anexo 14).

A diferencia de (Torres, Juan Carlos; Infante, Alfonso, 2011) quienes encuentran en su investigación que mientras mayor sea el ingreso de la familia, mayor es la posibilidad de contar con un computador y acceso a Internet.

En su estudio, realizado en Chile, (Ayala, 2007) encuentra que no existe relación significativa entre el nivel de ingresos con el logro académico, su hipótesis plantea que a mayor ingresos económicos, mejor logro académico.

Los experimentos realizados para responder a la **Hipótesis 2**, nos muestran que no existe una relación significativa entre el nivel de ingresos económicos y el uso de Internet en actividades de entretenimiento.

El análisis de los perfiles Medio, Bajo y Alto, nos muestra que en el perfil Alto (62,44%), es mayor la incidencia del uso de Redes Sociales, lo que convendría investigar es hasta qué punto se usa la red social como fuente de información o de uso para actividades académicas.

En el caso del perfil Medio (20,40%), se observa que la mayoría utiliza Youtube como fuente para su entretenimiento.

La hipótesis 3: el uso de Internet para actividades académicas y su influencia en el rendimiento académico no demostró una relación significativa, por lo que es evidente que el uso de Internet no define el rendimiento académico del estudiante.

La prueba χ^2 , para medir la diferencia entre los valores esperados y los valores observados, los resultados presentan un valor no significativo ($\chi^2 = 0.709$, $p = 0.950$; $p > 0.05$) y así se verifica que no existe incidencia significativa entre las variables dependientes e independientes, por lo que se rechaza H_1 .

Finalmente, en la **hipótesis 4**, el uso de Internet en actividades de entretenimiento no incide sobre el rendimiento académico. Para (Castaño M, 2011) el hecho de utilizar Internet con fines académicos no se traduce de forma automática en mejoras del rendimiento académico.

La prueba χ^2 , para medir la diferencia entre los valores esperados y los valores observados, los resultados presentan un valor no significativo ($\chi^2 = 0.332$, $p = 0.847$; $p > 0.05$) y así se verifica que no existe incidencia significativa entre las variables dependientes e independientes, por lo que se rechaza H_0 y podemos definir que no existe incidencia entre el uso de Internet en actividades de entretenimiento sobre el rendimiento académico (Anexo 16).

Sin embargo, podemos citar a (Ayala, 2007) que en su investigación encuentra diferencias en el uso de Internet de acuerdo a los diferentes grupos socioeconómicos, especialmente en el hecho que los estratos económicos, o aquellos que tienen menores ingresos, utilizan más Internet como fuente de entretenimiento, mientras que, aquellos que tienen más ingresos económicos utilizan Internet para acceder a servicios e información.

(López Jiménez & Bernal, 2011) en su investigación sobre el uso de Internet y entretenimiento realizada en Quito, muestra como aquéllos se entretienen más a diario y varias veces al día respecto a juegos en Internet, escuchar o bajar música, ver videos o bajarlos, escuchar radio, y visitar redes sociales, que los usuarios no estudiantes de Internet. En todos los casos es significativa la diferencia del consumo de hábitos de entretenimiento de los estudiantes, y en todos los casos hay una relación de causalidad.

CAPÍTULO VI: CONCLUSIONES Y RECOMENDACIONES

6.1 Conclusiones.

En este apartado en base a los resultados obtenidos se presenta las conclusiones obtenidas en el análisis y estudio del proyecto de investigación.

- Se demostró que el nivel de ingresos económicos mensuales de la familia de los estudiantes de la Escuela Superior Politécnica del Litoral (ESPOL) no influye en el uso de Internet en las actividades académicas, a pesar de que la mayor proporción de estudiantes sus familiares tienen ingresos entre 600 y 1000 dólares mensuales; este factor no ha sido un limitante para que los estudiantes puedan acceder a Internet desde sus hogares los 7 días de la semana; esto se puede corroborar al justificar el modelo de regresión logística binomial en el que los datos no se ajustaron al modelo propuesto.
- Se comprobó que el nivel de ingresos económicos no inciden en el uso de Internet para actividades de entretenimiento, al construir el modelo de regresión logística multinomial no se ajustó a los datos.
- El uso de Internet tanto para actividades académicas como de entretenimiento no incide sobre el rendimiento académico, puesto que el modelo de predicción de la regresión logística binomial arrojó como resultado que al ingresar la variable independiente sobre la variable dependiente el valor del suceso no varía; además, existe un alto porcentaje de alumnos que no han reprobado asignaturas.
- Al referirse a las técnicas empleadas para la generación de los modelos se demostró que la técnica más viable y utilizada fue la técnica de clustering, que junto al algoritmo de K-means permitió agrupar a las estudiantes de acuerdo a sus características comunes obteniendo resultados eficaces en cada grupo seleccionado.
- Se evidenció que, para la construcción del modelo de regresión logística se debe tener en cuenta el número de categorías contenidas en la variable dependiente; es decir, si ésta es dicotómica se utilizará la regresión logística binomial, en caso contrario, se empleará la regresión logística multinomial.

6.2 Recomendaciones y trabajos futuros.

Sobre la encuesta:

El diseño de la encuesta es de corte cuantitativo, permite medir el nivel de usos de Internet con fines académicos por parte de los estudiantes, sin embargo, se ha identificado limitaciones respecto al alcance de la medición del nivel de usos de Internet. Lo que nos permite identificar que el instrumento de recolección de información es perfectible acorde al tiempo y a las necesidades.

Sería interesante considerar que, a más de los lugares de conexión, se mida el tipo de conexión y la frecuencia realizada, por ejemplo, algunos estudiantes podrían conectarse a la red municipal de Internet, a través de los puntos de Internet Gratuito del Proyecto Guayaquil Digital⁷, otros a la red abierta de las universidades, uso de red celular (limitado o a través de un plan personal o corporativo).

Algunos ítems de la encuesta presentaron ambigüedad en los encuestados, lo que se identificó que contestaran erróneamente a lo solicitado, por ejemplo: Conocimientos en el manejo de Internet, años de conexión a Internet.

Si bien es cierto la encuesta fue diseñada en una escala tipo Likert, se podría pensar en reajustar el contenido de respuestas para que el encuestado tenga puntos o categoría de selección de la escala y de esta manera poder sumar las puntuaciones obtenidas en relación con todas las afirmaciones (Hernández Sampieri, Fernández Collado, & Baptista Lucio, 2010).

⁷ <http://guayaquil.gob.ec/content/puntos-de-internet-gratuito-wifi-guayaquildigital>

Brecha digital:

Realizar un análisis entre los niveles de uso de la Generación Z, la generación millenials y la disminución de la brecha digital.

Realizar una investigación sobre los niveles de uso de Internet en las Universidades públicas y de menor categoría.

Educational Data Mining:

La aplicación del KDD como parte de la Minería de Datos, permitió entender mejor el comportamiento de los datos y descubrir la influencia de los usos de Internet tanto para entretenimiento como para fines académicos en el presente trabajo de investigación.

El proceso de minería de datos y el apoyo de la Estadística como componente de la misma facilita el desarrollo de los procesos para analizar los datos con las herramientas estadísticas.

Promover centros de investigación sobre esta disciplina de la Minería de Datos y su alcance a nivel país.

Uso de Internet para actividades de entretenimiento:

Se podría realizar una investigación sobre los accesos a Internet desde consolas de video: PlayStation, Xbox, Nintendo. Así también el uso de Televisión por Internet como el caso de Netflix.

BIBLIOGRAFÍA

- Arnal, J., del Rincón, D., & Latorre, A. (1992). *Investigación educativa: fundamentos y metodología*. Barcelona: Labor.
- Avello Martínez, R., López Fernández, R., Cañedo Iglesias, M., Álvarez Acosta, H., Granados Romero, J., & Obando Freire, F. (2013). *Evolución de la alfabetización digital: nuevos conceptos y nuevas alfabetizaciones*. Obtenido de Medisur : <http://www.medisur.sld.cu/index.php/medisur/article/view/2467/1256>
- Ayala, C. (2007). Relación entre el Uso de Internet y el Logro Académico.
- Barnes, T., Desmarais, M., Romero, C., & Ventura, S. (2009). Educational Data Mining. "2nd International Conference on Educational Data Mining Proceedings". Córdoba (España).
- Caicedo, G. (23 de Abril de 2014). *Ecuador sigue disminuyendo la brecha digital*. Obtenido de <http://www.doctortecno.com/noticia/ecuador-sigue-disminuyendo-brecha-digital>
- Calders, T., & Pechenizkiy, M. (2012). *Introduction to the Special Section on Educational Data Mining*. Obtenido de <http://www.kdd.org/sites/default/files/issues/13-2-2011-12/V13-02-02-Calders%28introduction%29.pdf>
- Castaño M, J. (2011). *El uso de Internet para interacción en el aprendizaje: un análisis de la eficacia y la igualdad en el sistema universitario catalán*. Barcelona.
- Castaño M., J. (Enero de 2010). *La desigualdad digital entre los alumnos universitarios de los países desarrollados y su relación con el rendimiento académico*. Obtenido de http://rusc.uoc.edu/index.php/rusc/article/viewFile/v7n1_castano/v7n1_castano_esp
- Cobo, C. (Septiembre de 2011). *Aprendizaje y alfabetismo invisible ¿Por qué no basta con ser 2.0?* Obtenido de <https://ergonomic.wordpress.com/2011/09/07/no-basta-con-2-0/>
- Connolly, T. M., & Begg, C. E. (2005). *"Sistemas de bases de datos"*. Madrid: Pearson Educación S.A.
- de la Fuente Fernández, S. (2011). *Facultad de Ciencias Económicas y Empresariales*. Obtenido de Universidad Autónoma de Madrid:

<http://www.fuenterrebollo.com/Economicas/ECONOMETRIA/REDUCIR-DIMENSION/CORRESPONDENCIAS/correspondencias.pdf>

División de Datos y Estadísticas de TIC de la UIT. (2014). *Informe sobre Medición de la Sociedad de Información*. Ginebra: UIT.

Erazo, J. (26 de Abril de 2015). *Ecuador avanza en la reducción de la brecha digital*. Obtenido de <http://www.elciudadano.gob.ec/ecuador-avanza-en-la-reduccion-de-la-brecha-digital/>

Fernandes Fernandes, P. A. (s.f.). *Minería de Datos y Almacenamiento WEB*. Obtenido de <http://mineriadatosyalmacenamientoweb.net/datamining.html>

Galvis, M., & Martínez, F. (2004). *"Confrontación de dos técnicas de minería de datos aplicadas a un dominio específico"*. (Tesis de grado). Obtenido de www.javeriana.edu.co/biblos/tesis/ingenieria/Tesis184.pdf

Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techiques*. United States: Elsevier.

Hernández Orallo, J. (30 de Noviembre de 2006). *Universitat Politècnica de València*. Obtenido de <http://users.dsic.upv.es/~jorallo/master/dm2.pdf>

Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, P. (2010). *Metodología de la Investigación - Quinta edición*. México: McGraw-Hill.

INEC - Instituto Nacional de Estadística y Censos. (2012). *Tecnologías de la Información y Comunicación*. Obtenido de <https://www.youtube.com/watch?v=Lss5wgGyl8g>

INEC - Instituto Nacional de Estadística y Censos. (2013). *Tecnologías de la Información y Comunicaciones (TIC'S)*. Obtenido de http://www.ecuadorencifras.gob.ec/documentos/web-inec/Estadisticas_Sociales/TIC/Resultados_principales_140515.Tic.pdf

International Educational Data Mining Society. (2012). *International Educational Data Mining Society*. Obtenido de www.educationaldatamining.org

Lind, D. A., Marchal, W. G., & Wathen, S. A. (2012). *Estadística Aplicada a los Negocios y la Economía*. México: McGraw-Hill.

Martínez, E., & Serrano, A. (31 de Julio de 2007). *La Brecha Digital*. Obtenido de <http://www.labrechadigital.org/labrecha/Articulos/la-evolucion-hacia-una-nueva-brecha-digital.html>

- Martínez, E., Ascencio, I., & Serrano, A. (17 de Octubre de 2005). *La Brecha Digital*.
Obtenido de
<http://www.labrechadigital.org/labrecha/Articulos/understanding.html>
- Ministerio de Telecomunicaciones y Sociedad de la Información. (29 de Abril de 2014).
Ecuador sigue escalando posiciones en índice de reducción de brecha digital y uso de las TIC. Obtenido de <http://www.telecomunicaciones.gob.ec/ecuador-sigue-escalando-posiciones-en-indice-de-reduccion-de-brecha-digital-y-uso-de-las-tic/>
- Molina López, J. M., & García Herrero, J. (2006). *Técnicas de análisis de datos*.
- Molinari, A. (Agosto de 2011). *TED*. Obtenido de
http://www.ted.com/talks/aleph_molinari_let_s_bridge_the_digital_divide
- Peña, D. (23 de Enero de 2012). *Análisis de Datos Multivariantes*. Obtenido de
http://www.academia.edu/6134000/An%C3%A1lisis_de_Datos_Multivariantes_-_Daniel_Pe%C3%B1a
- Pérez López, C., & Santín González, D. (2007). *Minería de Datos: Técnicas y herramientas*. Madrid: Thomson Ediciones Paraninfo, S.A.
- Programa de las Naciones Unidas para el Desarrollo. (2014). *Informes sobre Desarrollo Humano 2014*. New York: PBM Graphics.
- Riquelme, J. C., Ruiz, R., & Gilbert, K. (2006). "*Minería de datos: Conceptos y tendencias*". Obtenido de <http://redalyc.uaemex.mx/pdf/925/92502902.pdf>
- Romero, C., & Ventura, S. (2007). "Educational Data Mining: a survey from 1995 to 2005". *Expert Systemas with Applications*, 135 - 146.
- Romero, C., & Ventura, S. (2010). "Educational Data Mining: A Review of the State-of-the-Art". *Transactions on Systems, Man, and Cybernetics*, 2.
- Romero, C., & Ventura, S. (2010). *Educational Data Mining: A review of the State-of-the-Art*. Obtenido de
<http://sci2s.ugr.es/keel/pdf/keel/articulo/VersionFinalEnviada.pdf>
- Romero, C., De Bra, P., & Ventura, S. (2004). "Knowledge discovery with genetic programming for providing feedback to courseware autor". *The Journal of Personalization Research*, (págs. 425 - 464).
- Ruiz, M., & Pardo, A. (24 de Enero de 2001). *Departamento de Sociología - Universidad Complutense*. Obtenido de

http://pendientedemigracion.ucm.es/info/socivmyt/paginas/D_departamento/materiales/analisis_datosyMultivariable/20factor_SPSS.pdf

Salazar Paredes, C. D. (2012). *Caracterización de deudores de tarjetas de crédito de una Institución financiera a través de técnicas de aprendizaje estadístico*.

Caracas: Universidad Central de Venezuela. Obtenido de

http://www.matematica.ciens.ucv.ve/modelos/tesis/MS_CarlosSalazar.pdf

Serrano Santoyo, A., & Martínez Martínez, E. (2003). *La brecha digital: mitos y realidades*. Nueva Mexicali: Editorial Universitaria - Universidad Autónoma de Baja California.

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". *Tesis Doctoral*.

U. Fayyad, e. a. (1996). *Advanced in Knowledge Discovery and Data Mining*. MA: MIT Press.

Universidad de Alicante. (11 de Abril de 2011). *Práctica 5. Análisis Multivariante con SPSS. Reducción de Datos. Análisis de componentes principales y factorial*.

Obtenido de <http://web.ua.es/es/lpa/docencia/practicas-analisis-exploratorio-de-datos-con-spss/practica-5-analisis-multivariante-con-spss-reduccion-de-datos-analisis-de-componentes-principales-y-factorial.html>

Villacís, D., & Jaramillo, A. (8 de Septiembre de 2014). El 20% de los ecuatorianos, es analfabeto digital. *El Comercio.com*. Obtenido de

<http://www.elcomercio.com.ec/tendencias/ecuatorianos-analfabeto-digital-cifras-tecnologia.html>

ANEXOS

Anexo 1: Instrumento de recolección de información.

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

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

2. Responda la siguiente pregunta	
¿Que carrera estudia?	

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

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

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

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

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

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

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

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

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

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

¿ Cuántos contactos tiene en LinkedIn?	(____)
--	----------

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

14. ¿Cuál es su nivel de uso de los siguientes dispositivos? (1 significa no usar y 10 significa utilizar al máximo)	1	2	3	4	5	6	7	8	9	10
Smartphone con cámara fotográfica y acceso a internet	0	0	0	0	0	0	0	0	0	0
Teléfono móvil con acceso a internet	0	0	0	0	0	0	0	0	0	0
Teléfono móvil sin acceso a internet	0	0	0	0	0	0	0	0	0	0
Computador portátil	0	0	0	0	0	0	0	0	0	0
Tablet (iPad; Galaxy Tab, Kindle, etc)	0	0	0	0	0	0	0	0	0	0
Cámara digital	0	0	0	0	0	0	0	0	0	0
iPod / MP3 Player	0	0	0	0	0	0	0	0	0	0

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	0	0	0	0	0	0	0	0	0	0
Usted confía en la información de internet para realizar sus tareas	0	0	0	0	0	0	0	0	0	0
Internet le permite prescindir de la Biblioteca	0	0	0	0	0	0	0	0	0	0
Internet facilita el proceso de aprendizaje	0	0	0	0	0	0	0	0	0	0
Internet le permite mejorar sus calificaciones	0	0	0	0	0	0	0	0	0	0
Usted presenta trabajos académicos copiados desde Internet	0	0	0	0	0	0	0	0	0	0

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: Análisis factorial de las variables relacionadas con actividades académicas de los estudiantes.

Matriz de correlacionesa											
	A_IngresoPlataforma	A_ConsultasProfesor	A_ConsultasEstudiantes	A_RecursosEducativos	A_VideosAcadémicos	A_ForosVirtuales	A_PostsAcadémicos	A_HorasChatAcadémico	A_HorasBúsqueda	A_BibliotecaVirtual	
Correlación	1.000	.177	.204	.276	.136	.087	.110	.253	.184	.187	
	A_IngresoPlataforma	1.000	.344	.186	.079	.175	.153	.252	.204	.120	
	A_ConsultasProfesor	.177	1.000	.344	.186	.079	.153	.252	.204	.120	
	A_ConsultasEstudiantes	.204	.344	1.000	.161	.125	.143	.283	.160	.125	
	A_RecursosEducativos	.276	.186	.161	1.000	.172	.114	.239	.240	.145	
	A_VideosAcadémicos	.136	.079	.125	.172	1.000	.189	.130	.201	.200	
	A_ForosVirtuales	.087	.175	.125	.172	.189	1.000	.207	.102	.158	
	A_PostsAcadémicos	.110	.153	.143	.114	.130	.207	1.000	.184	.069	
	A_HorasChatAcadémico	.253	.252	.283	.239	.201	.102	.184	1.000	.326	
	A_HorasBúsqueda	.184	.204	.160	.240	.200	.158	.069	.326	1.000	
	A_BibliotecaVirtual	.187	.120	.125	.145	.063	.156	.161	.228	.167	
Sig. (Unilateral)	A_IngresoPlataforma	.000	.000	.000	.000	.001	.028	.008	.000	.000	
	A_ConsultasProfesor	.000	.000	.000	.000	.040	.000	.000	.000	.004	
	A_ConsultasEstudiantes	.000	.000	.000	.000	.003	.003	.001	.000	.003	
	A_RecursosEducativos	.000	.000	.000	.000	.000	.006	.000	.000	.001	
	A_VideosAcadémicos	.001	.040	.003	.000	.000	.002	.000	.000	.082	
	A_ForosVirtuales	.028	.000	.003	.000	.000	.000	.012	.000	.000	
	A_PostsAcadémicos	.008	.000	.001	.006	.002	.000	.000	.065	.000	
	A_HorasChatAcadémico	.000	.000	.000	.000	.000	.012	.000	.000	.000	
	A_HorasBúsqueda	.000	.000	.000	.000	.000	.000	.065	.000	.000	
	A_BibliotecaVirtual	.000	.004	.003	.001	.082	.000	.000	.000	.000	

a. Determinante = .370

Varianza total explicada						
Componente	Autovalores iniciales			extracción		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	2.616	26.164	26.164	2.616	26.164	26.164
2	1.058	10.576	36.740	1.058	10.576	36.740
3	1.018	10.177	46.917	1.018	10.177	46.917
4	.953	9.525	56.442			
5	.867	8.667	65.109			
6	.843	8.429	73.538			
7	.750	7.502	81.040			
8	.667	6.673	87.713			
9	.631	6.313	94.026			
10	.597	5.974	100.000			

KMO y prueba de Bartlett		
Medida de adecuación		.783
Prueba de esfericidad de Bartlett	Chi-cuadrado	478.126
	gl	45
	Sig.	.000

Anexo 3: Análisis factorial de las variables relacionadas con actividades de entretenimiento de los estudiantes.

Matriz de correlacionesa						
		E_HorasChatP ersonal	E_RedesSocia les	E_JuegosLíne a	E_DescargaM úsica	E_Videos
Correlación	E_HorasChatP ersonal	1.000	.504	.207	.281	.192
	E_RedesSocia les	.504	1.000	.103	.235	.236
	E_JuegosLíne	.207	.103	1.000	.254	.239
	E_DescargaM	.281	.235	.254	1.000	.216
	E_Videos	.192	.236	.239	.216	1.000
Sig. (Unilateral)	E_HorasChatP		.000	.000	.000	.000
	E_RedesSocia	.000		.011	.000	.000
	E_JuegosLíne	.000	.011		.000	.000
	E_DescargaM	.000	.000	.000		.000
	E_Videos	.000	.000	.000	.000	

a. Determinante = 0.548

Varianza total explicada						
Componente	Autovalores iniciales			extracción		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	2.005	40.093	40.093	2.005	40.093	40.093
2	1.004	20.080	60.174	1.004	20.080	60.174
3	.797	15.945	76.119			
4	.719	14.380	90.499			
5	.475	9.501	100.000			

KMO y prueba de Bartlett		
Medida de adecuación		.656
Prueba de esfericidad de Bartlett	Chi-cuadrado aproximado	290.119
	gl	10
	Sig.	.000

Anexo 4: Análisis del Clúster en dos grupos según el uso de Internet en actividades académicas de los estudiantes.

Centros de los conglomerados finales

	Conglomerado	
	1	2
¿Aproximadamente cuántas horas busca información académica en internet cada mes?	14	38
¿Aproximadamente cuántas horas chatea sobre temas académicos cada mes?	8	9
¿Aproximadamente cuántas consultas realiza a sus compañeros por mes?	10	11
¿Aproximadamente cuántos recursos educativos descarga de la plataforma virtual cada mes?	7	9
¿Aproximadamente cuántos videos académicos mira en youtube cada mes?	7	8

Distancias entre los centros de los conglomerados finales

Conglomerado	1	2
1		24,982
2	24,982	

ANOVA

	Conglomerado		Error		F	Sig.
	Media cuadrática	gl	Media cuadrática	gl		
¿Aproximadamente cuántas horas busca información académica en internet cada mes?	58477,937	1	62,812	488	931,002	,000
¿Aproximadamente cuántas horas chatea sobre temas académicos cada mes?	327,277	1	32,840	488	9,966	,002
¿Aproximadamente cuántas consultas realiza a sus compañeros por mes?	108,339	1	74,950	488	1,445	,230
¿Aproximadamente cuántos recursos educativos descarga de la plataforma virtual cada mes?	254,372	1	30,350	488	8,381	,004
¿Aproximadamente cuántos videos académicos mira en youtube cada mes?	144,896	1	32,109	488	4,513	,034

Las pruebas F sólo se deben utilizar con una finalidad descriptiva puesto que los conglomerados han sido elegidos para maximizar las diferencias entre los casos en diferentes conglomerados. Los niveles críticos no son corregidos, por lo que no pueden interpretarse como pruebas de la hipótesis de que los centros de los conglomerados son iguales.

Número de casos en cada conglomerado

Conglomerado	1	361,000
	2	129,000
Válidos		490,000
Perdidos		,000

Anexo 5: Análisis del Clúster en tres grupos según el uso de Internet en actividades académicas de los estudiantes.

Centros de los conglomerados finales

	Conglomerado		
	1	2	3
¿Aproximadamente cuántas horas busca información académica en internet cada mes?	10	21	41
¿Aproximadamente cuántas horas chatea sobre temas académicos cada mes?	5	11	9
¿Aproximadamente cuántas consultas realiza a sus compañeros por mes?	5	17	8
¿Aproximadamente cuántos recursos educativos descarga de la plataforma virtual cada mes?	5	10	9
¿Aproximadamente cuántos videos académicos mira en youtube cada mes?	6	9	8

Distancias entre los centros de los conglomerados finales

Conglomerado	1	2	3
1		18,485	31,828
2	18,485		22,286
3	31,828	22,286	

ANOVA

	Conglomerado		Error		F	Sig.
	Media cuadrática	gl	Media cuadrática	gl		
¿Aproximadamente cuántas horas busca información académica en internet cada mes?	33025,669	2	47,390	487	696,896	,000
¿Aproximadamente cuántas horas chatea sobre temas académicos cada mes?	2008,602	2	25,331	487	79,295	,000
¿Aproximadamente cuántas consultas realiza a sus compañeros por mes?	7500,272	2	44,525	487	168,452	,000
¿Aproximadamente cuántos recursos educativos descarga de la plataforma virtual cada mes?	1128,018	2	26,302	487	42,886	,000
¿Aproximadamente cuántos videos académicos mira en youtube cada mes?	388,167	2	30,878	487	12,571	,000

Las pruebas F sólo se deben utilizar con una finalidad descriptiva puesto que los conglomerados han sido elegidos para maximizar las diferencias entre los casos en diferentes conglomerados. Los niveles críticos no son corregidos, por lo que no pueden interpretarse como pruebas de la hipótesis de que los centros de los conglomerados son iguales.

Número de casos en cada conglomerado

Conglomerado	1	211,000
	2	179,000
	3	100,000
Válidos		490,000
Perdidos		,000

Anexo 6: Análisis del Clúster en cuatro grupos según el uso de Internet en actividades académicas de los estudiantes.

Centros de los conglomerados finales

	Conglomerado			
	1	2	3	4
¿Aproximadamente cuántas horas busca información académica en internet cada mes?	48	7	23	20
¿Aproximadamente cuántas horas chatea sobre temas académicos cada mes?	10	4	9	11
¿Aproximadamente cuántas consultas realiza a sus compañeros por mes?	10	5	6	22
¿Aproximadamente cuántos recursos educativos descarga de la plataforma virtual cada mes?	9	5	9	9
¿Aproximadamente cuántos videos académicos mira en youtube cada mes?	10	5	8	8

Distancias entre los centros de los conglomerados finales

Conglomerado	1	2	3	4
1		41,391	24,940	30,055
2	41,391		17,202	22,956
3	24,940	17,202		16,696
4	30,055	22,956	16,696	

ANOVA

	Conglomerado		Error		F	Sig.
	Media cuadrática	gl	Media cuadrática	gl		
¿Aproximadamente cuántas horas busca información académica en internet cada mes?	24944,667	3	29,416	486	848,001	,000
¿Aproximadamente cuántas horas chatea sobre temas académicos cada mes?	1319,520	3	25,504	486	51,739	,000
¿Aproximadamente cuántas consultas realiza a sus compañeros por mes?	7778,738	3	27,465	486	283,226	,000
¿Aproximadamente cuántos recursos educativos descarga de la plataforma virtual cada mes?	646,798	3	27,006	486	23,950	,000
¿Aproximadamente cuántos videos académicos mira en youtube cada mes?	354,141	3	30,353	486	11,667	,000

Las pruebas F sólo se deben utilizar con una finalidad descriptiva puesto que los conglomerados han sido elegidos para maximizar las diferencias entre los casos en diferentes conglomerados. Los niveles críticos no son corregidos, por lo que no pueden interpretarse como pruebas de la hipótesis de que los centros de los conglomerados son iguales.

Número de casos en cada conglomerado

Conglomerado	1	60,000
	2	171,000
	3	143,000
	4	116,000
Válidos		490,000
Perdidos		,000

Anexo 7: Análisis del Clúster en cinco grupos según el uso de Internet en actividades académicas de los estudiantes.

Centros de los conglomerados finales

	Conglomerado				
	1	2	3	4	5
¿Aproximadamente cuántas horas busca información académica en internet cada mes?	20	23	43	49	7
¿Aproximadamente cuántas horas chatea sobre temas académicos cada mes?	12	9	7	11	4
¿Aproximadamente cuántas consultas realiza a sus compañeros por mes?	22	6	20	6	5
¿Aproximadamente cuántos recursos educativos descarga de la plataforma virtual cada mes?	9	9	8	10	5
¿Aproximadamente cuántos videos académicos mira en youtube cada mes?	8	8	8	10	5

Distancias entre los centros de los conglomerados finales

Conglomerado	1	2	3	4	5
1		16,834	23,836	32,866	22,876
2	16,834		24,201	25,629	17,195
3	23,836	24,201		15,188	38,612
4	32,866	25,629	15,188		42,261
5	22,876	17,195	38,612	42,261	

ANOVA

	Conglomerado		Error		F	Sig.
	Media cuadrática	gl	Media cuadrática	gl		
¿Aproximadamente cuántas horas busca información académica en internet cada mes?	18763,832	4	29,020	485	646,580	,000
¿Aproximadamente cuántas horas chatea sobre temas académicos cada mes?	1096,302	4	24,676	485	44,427	,000
¿Aproximadamente cuántas consultas realiza a sus compañeros por mes?	6291,358	4	23,750	485	264,901	,000
¿Aproximadamente cuántos recursos educativos descarga de la plataforma virtual cada mes?	493,402	4	26,993	485	18,279	,000
¿Aproximadamente cuántos videos académicos mira en youtube cada mes?	271,200	4	30,370	485	8,930	,000

Las pruebas F sólo se deben utilizar con una finalidad descriptiva puesto que los conglomerados han sido elegidos para maximizar las diferencias entre los casos en diferentes conglomerados. Los niveles críticos no son corregidos, por lo que no pueden interpretarse como pruebas de la hipótesis de que los centros de los conglomerados son iguales.

Número de casos en cada conglomerado

Conglomerado	1	111,000
	2	144,000
	3	21,000
	4	43,000
	5	171,000
Válidos		490,000
Perdidos		,000

Anexo 8: Análisis Discriminante para probar el porcentaje de exactitud de actividades académicas.

Resultados de la clasificación^a

Número inicial de casos			Grupo de pertenencia pronosticado					Total
			1	2	3	4	5	
Original	Recuento	1	108	2	0	0	1	111
		2	0	142	0	0	2	144
		3	1	0	18	2	0	21
		4	0	0	0	43	0	43
		5	2	1	0	0	168	171
%		1	97,3	1,8	,0	,0	,9	100,0
		2	,0	98,6	,0	,0	1,4	100,0
		3	4,8	,0	85,7	9,5	,0	100,0
		4	,0	,0	,0	100,0	,0	100,0
		5	1,2	,6	,0	,0	98,2	100,0

a. Clasificados correctamente el 97.8% de los casos agrupados originales.

Resultados de la clasificación^a

Número inicial de casos			Grupo de pertenencia pronosticado				Total
			1	2	3	4	
Original	Recuento	1	108	3	0	0	111
		2	0	144	0	0	144
		3	3	0	16	2	21
		4	0	0	0	43	43
		Casos desagrupados	12	159	0	0	171
%		1	97,3	2,7	,0	,0	100,0
		2	,0	100,0	,0	,0	100,0
		3	14,3	,0	76,2	9,5	100,0
		4	,0	,0	,0	100,0	100,0
		Casos desagrupados	7,0	93,0	,0	,0	100,0

a. Clasificados correctamente el 97.5% de los casos agrupados originales.

Resultados de la clasificación^a

Número inicial de casos			Grupo de pertenencia pronosticado			Total
			1	2	3	
Original	Recuento	1	107	4	0	111
		2	0	144	0	144
		3	3	0	18	21
		Casos desagrupados	11	171	32	214
%		1	96,4	3,6	,0	100,0
		2	,0	100,0	,0	100,0
		3	14,3	,0	85,7	100,0
		Casos desagrupados	5,1	79,9	15,0	100,0

a. Clasificados correctamente el 97.5% de los casos agrupados originales.

Resultados de la clasificación^a

Número inicial de casos			Grupo de pertenencia pronosticado		Total
			1	2	
Original	Recuento	1	107	4	111
		2	0	144	144
		Casos desagrupados	24	211	235
%		1	96,4	3,6	100,0
		2	,0	100,0	100,0
		Casos desagrupados	10,2	89,8	100,0

a. Clasificados correctamente el 98.4% de los casos agrupados originales.

Anexo 9: Análisis del Clúster en dos grupos según el uso de Internet en actividades de entretenimiento.

Centros de los conglomerados finales

	Conglomerado	
	1	2
¿ Aproximadamente cuántas horas a la semana chatea por diversión?	7	17
¿Aproximadamente cuántas horas a la semana utiliza redes sociales?	8	22
¿Aproximadamente cuántos videos para entretenimiento mira en youtube cada semana?	7	14

Distancias entre los centros de los conglomerados finales

Conglomerado	1	2
1		18,921
2	18,921	

ANOVA

	Conglomerado		Error		F	Sig.
	Media cuadrática	gl	Media cuadrática	gl		
¿ Aproximadamente cuántas horas a la semana chatea por diversión?	9955,491	1	34,694	488	286,951	,000
¿Aproximadamente cuántas horas a la semana utiliza redes sociales?	21337,983	1	35,005	488	609,565	,000
¿Aproximadamente cuántos videos para entretenimiento mira en youtube cada semana?	5108,213	1	49,589	488	103,012	,000

Las pruebas F sólo se deben utilizar con una finalidad descriptiva puesto que los conglomerados han sido elegidos para maximizar las diferencias entre los casos en diferentes conglomerados. Los niveles críticos no son corregidos, por lo que no pueden interpretarse como pruebas de la hipótesis de que los centros de los conglomerados son iguales.

Número de casos en cada conglomerado

Conglomerado	1	346,000
	2	144,000
Válidos		490,000
Perdidos		,000

Anexo 10: Análisis del Clúster en tres grupos según el uso de Internet en actividades de entretenimiento.

Centros de los conglomerados finales

	Conglomerado		
	1	2	3
¿ Aproximadamente cuántas horas a la semana chatea por diversión?	7	19	10
¿Aproximadamente cuántas horas a la semana utiliza redes sociales?	8	25	12
¿Aproximadamente cuántos videos para entretenimiento mira en youtube cada semana?	5	10	22

ANOVA

	Conglomerado		Error		F	Sig.
	Media cuadrática	gl	Media cuadrática	gl		
¿ Aproximadamente cuántas horas a la semana chatea por diversión?	4848,052	2	35,298	487	137,347	,000
¿Aproximadamente cuántas horas a la semana utiliza redes sociales?	11798,067	2	30,440	487	387,580	,000
¿Aproximadamente cuántos videos para entretenimiento mira en youtube cada semana?	9593,130	2	20,783	487	461,590	,000

Las pruebas F sólo se deben utilizar con una finalidad descriptiva puesto que los conglomerados han sido elegidos para maximizar las diferencias entre los casos en diferentes conglomerados. Los niveles críticos no son corregidos, por lo que no pueden interpretarse como pruebas de la hipótesis de que los centros de los conglomerados son iguales.

**Número de casos en cada
conglomerado**

Conglomerado	1	306,000
	2	100,000
	3	84,000
Válidos		490,000
Perdidos		,000

**Distancias entre los centros de los conglomerados
finales**

Conglomerado	1	2	3
1		21,539	17,729
2	21,539		20,233
3	17,729	20,233	

Anexo 11: Análisis del Clúster en cuatro grupos según el uso de Internet en actividades de entretenimiento.

Centros de los conglomerados finales

	Conglomerado				
	1	2	3	4	5
¿ Aproximadamente cuántas horas a la semana chatea por diversión?	6	18	9	25	9
¿Aproximadamente cuántas horas a la semana utiliza redes sociales?	7	12	28	26	11
¿Aproximadamente cuántos videos para entretenimiento mira en youtube cada semana?	5	7	9	12	23

Distancias entre los centros de los conglomerados finales

Conglomerado	1	2	3	4
1		23,034	18,095	19,999
2	23,034		22,183	16,116
3	18,095	22,183		19,840
4	19,999	16,116	19,840	

ANOVA

	Conglomerado		Error		F	Sig.
	Media cuadrática	gl	Media cuadrática	gl		
¿ Aproximadamente cuántas horas a la semana chatea por diversión?	5367,369	3	22,190	486	241,888	,000
¿Aproximadamente cuántas horas a la semana utiliza redes sociales?	8289,309	3	27,886	486	297,256	,000
¿Aproximadamente cuántos videos para entretenimiento mira en youtube cada semana?	6099,871	3	22,650	486	269,311	,000

Las pruebas F sólo se deben utilizar con una finalidad descriptiva puesto que los conglomerados han sido elegidos para maximizar las diferencias entre los casos en diferentes conglomerados. Los niveles críticos no son corregidos, por lo que no pueden interpretarse como pruebas de la hipótesis de que los centros de los conglomerados son iguales.

Número de casos en cada conglomerado

Conglomerado	1	284,000
	2	51,000
	3	76,000
	4	79,000
Válidos		490,000
Perdidos		,000

Anexo 12: Análisis del Clúster en cinco grupos según el uso de Internet en actividades de entretenimiento.

Centros de los conglomerados finales

	Conglomerado				
	1	2	3	4	5
¿ Aproximadamente cuántas horas a la semana chatea por diversión?	6	18	9	25	9
¿Aproximadamente cuántas horas a la semana utiliza redes sociales?	7	12	28	26	11
¿Aproximadamente cuántos videos para entretenimiento mira en youtube cada semana?	5	7	9	12	23

Distancias entre los centros de los conglomerados finales

Conglomerado	1	2	3	4	5
1		13,511	22,224	28,272	18,446
2	13,511		18,168	16,257	17,538
3	22,224	18,168		16,194	21,565
4	28,272	16,257	16,194		24,404
5	18,446	17,538	21,565	24,404	

ANOVA

	Conglomerado		Error		F	Sig.
	Media cuadrática	gl	Media cuadrática	gl		
¿ Aproximadamente cuántas horas a la semana chatea por diversión?	4490,819	4	18,398	485	244,095	,000
¿Aproximadamente cuántas horas a la semana utiliza redes sociales?	6887,560	4	22,413	485	307,301	,000
¿Aproximadamente cuántos videos para entretenimiento mira en youtube cada semana?	4660,485	4	21,991	485	211,929	,000

Las pruebas F sólo se deben utilizar con una finalidad descriptiva puesto que los conglomerados han sido elegidos para maximizar las diferencias entre los casos en diferentes conglomerados. Los niveles críticos no son corregidos, por lo que no pueden interpretarse como pruebas de la hipótesis de que los centros de los conglomerados son iguales.

Número de casos en cada conglomerado

Conglomerado	1	258,000
	2	70,000
	3	46,000
	4	40,000
	5	76,000
Válidos		490,000
Perdidos		,000

Anexo 13: Análisis Discriminante para probar el porcentaje de exactitud de actividades de entretenimiento.

Resultados de la clasificación^a

Número inicial de casos			Grupo de pertenencia pronosticado					Total
			1	2	3	4	5	
Original	Recuento	1	253	3	1	0	1	258
		2	6	63	0	1	0	70
		3	0	0	46	0	0	46
		4	0	0	1	39	0	40
		5	0	1	1	0	74	76
%		1	98,1	1,2	,4	,0	,4	100,0
		2	8,6	90,0	,0	1,4	,0	100,0
		3	,0	,0	100,0	,0	,0	100,0
		4	,0	,0	2,5	97,5	,0	100,0
		5	,0	1,3	1,3	,0	97,4	100,0

a. Clasificados correctamente el 96.9% de los casos agrupados originales.

Resultados de la clasificación^a

Número inicial de casos			Grupo de pertenencia pronosticado				Total
			1	2	3	4	
Original	Recuento	1	253	4	1	0	258
		2	6	63	0	1	70
		3	0	0	46	0	46
		4	0	0	1	39	40
		Casos desagrupados	54	14	7	1	76
%		1	98,1	1,6	,4	,0	100,0
		2	8,6	90,0	,0	1,4	100,0
		3	,0	,0	100,0	,0	100,0
		4	,0	,0	2,5	97,5	100,0
		Casos desagrupados	71,1	18,4	9,2	1,3	100,0

a. Clasificados correctamente el 96.9% de los casos agrupados originales.

Resultados de la clasificación^a

Número inicial de casos			Grupo de pertenencia pronosticado			Total
			1	2	3	
Original	Recuento	1	253	4	1	258
		2	6	64	0	70
		3	0	0	46	46
		Casos desagrupados	56	46	14	116
%		1	98,1	1,6	,4	100,0
		2	8,6	91,4	,0	100,0
		3	,0	,0	100,0	100,0
		Casos desagrupados	48,3	39,7	12,1	100,0

a. Clasificados correctamente el 97.1% de los casos agrupados originales.

Resultados de la clasificación^a

Número inicial de casos			Grupo de pertenencia pronosticado		Total
			1	2	
Original	Recuento	1	253	5	258
		2	6	64	70
		Casos desagrupados	70	92	162
%		1	98,1	1,9	100,0
		2	8,6	91,4	100,0
		Casos desagrupados	43,2	56,8	100,0

a. Clasificados correctamente el 96.6% de los casos agrupados originales.

Anexo 14: Regresión Logística Binaria aplicada al perfil de Uso de Internet y actividades académicas.

Codificación de la variable dependiente	
Valor original	Valor interno
Tradicional	0
Tecnológico	1

Codificaciones de variables categóricas						
		Frecuencia	Codificación de parámetros			
			(1)	(2)	(3)	(4)
Los ingresos mensuales de su familia son de:	Hasta 350 dólares	84	.000	.000	.000	.000
	Hasta 600 dólares	144	1.000	.000	.000	.000
	Hasta 1000 dólares	125	.000	1.000	.000	.000
	Hasta 1.500 dólares	75	.000	.000	1.000	.000
	Más de 1.500 dólares	62	.000	.000	.000	1.000

Pruebas omnibus sobre los coeficientes del modelo				
		Chi cuadrado	gl	Sig.
Paso 1	Paso	.714	4	.950
	Bloque	.714	4	.950
	Modelo	.714	4	.950

Resumen del modelo			
Paso	-2 log de la verosimilitud	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	564.202	.001	.002

Prueba de Hosmer y Lemeshow			
Paso	Chi cuadrado	gl	Sig.
1	.000	3	1.000

Tabla de contingencias para la prueba de Hosmer y Lemeshow						
		binomial aprendizaje = .00		binomial aprendizaje = 1.00		Total
		Observado	Esperado	Observado	Esperado	
Paso 1	1	35	35.000	90	90.000	125
	2	21	21.000	54	54.000	75
	3	38	38.000	106	106.000	144
	4	15	15.000	47	47.000	62
	5	20	20.000	64	64.000	84

Pruebas de chi-cuadrado			
	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	.709	4	.950
Razón de verosimilitudes	.714	4	.950
Asociación lineal por lineal	.050	1	.823
N de casos válidos	490		

a. 0 casillas (.0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 16.32.

Anexo 15: Regresión Logística Multinomial aplicada al perfil de Uso de Internet y actividades de entretenimiento.

Información del ajuste del modelo						
Modelo	Criterio de ajuste del modelo			Contrastes de la razón de verosimilitud		
	AIC	BIC	-2 log verosimilitud	Chi-cuadrado	gl	Sig.
Sólo la intersección	59.557	67.946	55.557			
Final	64.470	106.414	44.470	11.088	8	.197

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

Pseudo R-cuadrado	
Cox y Snell	.022
Nagelkerke	.027
McFadden	.012

Contrastes de la razón de verosimilitud						
Efecto	Criterio de ajuste del modelo			Contrastes de la razón de verosimilitud		
	AIC de modelo reducido	BIC de modelo reducido	-2 log verosimilitud del modelo reducido	Chi-cuadrado	gl	Sig.
Intersección	64.470	106.414	44.470	.000	0	.
Ingresos	59.557	67.946	55.557	11.088	8	.197

El estadístico de chi-cuadrado es la diferencia en las -2 log verosimilitudes 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 ese efecto son 0.

a. Este modelo reducido es equivalente al modelo final ya que la omisión del efecto no incrementa los grados de libertad.

Anexo 16: Regresión Logística Binaria aplicada al Uso de Internet y actividades académicas sobre el rendimiento académico.

Codificación de la variable dependiente	
Valor original	Valor interno
Reprobado	0
Aprobado	1

Codificaciones de variables categóricas			
		Frecuencia	Codificación de parámetros (1)
Número inicial de casos	Tecnológico	361	1.000
	Tradicional	129	.000

Pruebas omnibus sobre los coeficientes del modelo				
		Chi cuadrado	gl	Sig.
Paso 1	Paso	.584	1	.445
	Bloque	.584	1	.445
	Modelo	.584	1	.445

Resumen del modelo			
Paso	-2 log de la verosimilitud	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	662.075	.001	.002

a. La estimación ha finalizado en el número de iteración 3 porque las estimaciones de los parámetros han cambiado en menos de .001.

Prueba de Hosmer y Lemeshow			
Paso	Chi cuadrado	gl	Sig.
1	.000	0	.

Tabla de contingencias para la prueba de Hosmer y Lemeshow						
		estudiantes = 0		estudiantes = Aprobado		Total
		Observado	Esperado	Observado	Esperado	
Paso 1	1	151	151.000	210	210.000	361
	2	49	49.000	80	80.000	129

Pruebas de chi-cuadrado					
	Valor	gl	Sig. asintótica (bilateral)	Sig. exacta (bilateral)	Sig. exacta (unilateral)
Chi-cuadrado de Pearson	.581	1	.446		
Corrección por continuidad ^b	.433	1	.511		
Razón de verosimilitudes	.584	1	.445		
Estadístico exacto de Fisher				.467	.256
Asociación lineal por lineal	.580	1	.446		
N de casos válidos	490				

a. 0 casillas (.0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 52.65.

b. Calculado sólo para una tabla de 2x2.

Anexo 17: Regresión Logística Binaria aplicada al Uso de Internet y actividades de entretenimiento sobre el rendimiento académico.

Codificación de la variable dependiente	
Valor original	Valor interno
Reprobado	0
Aprobado	1

Pruebas omnibus sobre los coeficientes del modelo				
		Chi cuadrado	gl	Sig.
Paso 1	Paso	.165	1	.685
	Bloque	.165	1	.685
	Modelo	.165	1	.685

Resumen del modelo			
Paso	-2 log de la verosimilitud	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	662.495	.000	.000

a. La estimación ha finalizado en el número de iteración 3 porque las estimaciones de los parámetros han cambiado en menos de .001.

Prueba de Hosmer y Lemeshow			
Paso	Chi cuadrado	gl	Sig.
1	.170	1	.681

Tabla de contingencias para la prueba de Hosmer y Lemeshow						
		estudiantes = 0		estudiantes = Aprobado		Total
		Observado	Esperado	Observado	Esperado	
Paso 1	1	126	126.859	180	179.141	306
	2	42	40.282	58	59.718	100
	3	32	32.859	52	51.141	84

Pruebas de chi-cuadrado			
	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	.332	2	.847
Razón de verosimilitudes	.334	2	.846
Asociación lineal por lineal	.164	1	.686
N de casos válidos	490		

a. 0 casillas (.0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 34.29.