



UNIVERSIDAD TÉCNICA PARTICULAR DE LOJA
La Universidad Católica de Loja

ÁREA TÉCNICA

**TITULO DE INGENIERO EN SISTEMAS INFORMÁTICOS Y
COMPUTACIÓN**

**Análisis de los Usos de la Tecnología en los Estudiantes de la Universidad
Técnica Particular de Loja.**

TRABAJO DE TITULACIÓN.

AUTORA: Veintimilla Toro, Irina Anabella.

DIRECTORA: Arias Tapia, Susana Alexandra, Ms.

**LOJA-ECUADOR
2015**



Esta versión digital, ha sido acreditada bajo la licencia Creative Commons 4.0, CC BY-NY-SA: Reconocimiento-No comercial-Compartir igual; la cual permite copiar, distribuir y comunicar públicamente la obra, mientras se reconozca la autoría original, no se utilice con fines comerciales y se permiten obras derivadas, siempre que mantenga la misma licencia al ser divulgada. <http://creativecommons.org/licenses/by-nc-sa/4.0/deed.es>

Septiembre, 2015

APROBACIÓN DE LA DIRECTORA DEL TRABAJO DE TITULACIÓN

Magister.

Susana Alexandra Arias Tapia

DOCENTE DE LA TITULACIÓN

De mi consideración:

Que el presente trabajo de titulación: Análisis de los Usos de la Tecnología en los Estudiantes de la Universidad Técnica Particular De Loja, realizado por Veintimilla Toro Irina Anabella, ha sido orientado y revisado durante su ejecución, por lo que se aprueba la presentación del mismo.

Loja, Noviembre de 2015

f).

DECLARACIÓN DE AUTORÍA Y CESIÓN DE DERECHOS

“Yo, Irina Anabella Veintimilla Toro, declaro ser autora del presente trabajo de titulación: Análisis de los Usos de la Tecnología en los Estudiantes de la Universidad Técnica Particular de Loja, de la Titulación de Sistemas Informáticos y Computación, siendo Susana Alexandra Arias Tapia, directora del presente trabajo; y eximo expresamente a la Universidad Técnica Particular de Loja y a sus representantes legales de posibles reclamos o acciones legales. Además certifico que las ideas, conceptos, procedimientos y resultados vertidos en el presente trabajo investigativo, son de mi exclusiva responsabilidad.

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

F.....
Veintimilla Toro Irina Anabella
0704952738

DEDICATORIA

Con mucho amor y cariño dedico el presente proyecto a:

Dios, por siempre acompañarme a lo largo de mi vida, por haberme dado la fe, sabiduría fortaleza y salud, para lograr mis objetivos,

Mis padres, Jaime e Ideni quienes desde pequeña me enseñaron a luchar para alcanzar mis objetivos. Todo lo que soy se lo debo a ellos, este triunfo es de ustedes papitos, los amo.

Mis hermanos, Dayana y Ronny por siempre darme su fuerza, apoyo incondicional, y también que con sus ocurrencias alegran mi vida.

Mi amado, Mario Andrés compañero inseparable de vida por brindarme cada día, amor, dulzura y su apoyo constante,

Mi pequeño hijo Dylan Andrés, que es la razón de mi existencia, ha llenado de alegría a mi corazón, y ha sido mi motivación principal para conseguir este logro tan anhelado,

A todos mis compañeros y amigos por brindarme su amistad sincera y estar presentes en los en los malos y buenos momentos

Irina Anabella

AGRADECIMIENTO

Este proyecto de titulación, es el resultado de mucho esfuerzo y dedicación, para culminarlo con éxito, implicó un apoyo constante de personas que siempre estuvieron de una u otra forma colaborando o participando para la realización del mismo, es por esto que debo agradecer:

En primer lugar, a Dios y la Virgen Santísima “María Auxiliadora” por haberme guiado por el camino de la felicidad, brindándome la sabiduría y la salud para la culminación de este trabajo.

A mis padres Jaime e Ideni, quienes siempre a pesar de las adversidades de la vida, con su inalcanzable esfuerzo supieron darme una carrera para mi futuro, creyendo en mí, sin dudar en ningún momento de mis habilidades.

A mi familia, mis hermanos, tíos, primos, amigas, compañeros y profesores por su apoyo y comprensión en todo momento a lo largo de mi vida académica, y por ser partícipes en este nuevo logro de mi vida.

En segundo lugar, merecen un agradecimiento especial mi Directora de tesis Ms. Susana Alexandra Arias Tapia, que con sus directrices, sus consejos, su esfuerzo, su paciencia y su conocimiento me guió en el desarrollo de este trabajo.

Al Dr. Juan Carlos Torres Días, por confiar en mí y en mi crecimiento profesional, brindándome la oportunidad de participar y desarrollar este proyecto,

A mi querido Suegro Magister. Mario Rigoberto Paz Ocampos y a su Señora esposa por siempre brindarme su apoyo incondicional.

Irina Anabella

ÍNDICE

PORTADA	i
APROBACIÓN DEL DIRECTOR DEL TRABAJO DE TITULACIÓN.....	ii
DECLARACIÓN DE AUTORÍA Y CESIÓN DE DERECHOS.....	iii
DEDICATORIA	iv
AGRADECIMIENTO	v
ÍNDICE	vi
ÍNDICE DE TABLAS.....	ix
ÍNDICE DE FIGURAS.....	xi
TERMINOLOGÍA	xiii
ÍNDICE DE ANEXOS.....	xiv
RESUMEN.....	1
ABSTRACT	2
INTRODUCCIÓN.....	3
CAPITULO I: PLANTEAMIENTO DE LA INVESTIGACIÓN	5
1.1. Objetivos	6
1.2. Preguntas de Investigación.....	6
1.3. Hipótesis.....	6
1.4. Procedimiento.....	6
CAPÍTULO II: MARCO TEÓRICO Y CONCEPTUAL	8
2.1. Importancia de las Tecnologías de Información y Comunicación.....	9
2.2. Brecha Digital	9
2.2.1. Introducción.....	9
2.2.2. Definiciones.....	10
2.2.3. Aspectos de la Brecha Digital	11
2.2.3.1. Brecha Digital de Acceso	12
2.2.3.2. Brecha Digital de Uso	12
2.2.3.3. Brecha Digital de Calidad de uso	12
2.2.4. Factores que determinan la Brecha Digital.	13
2.3. Uso de Internet.....	15
2.3.1. Uso y gratificación	15

2.3.2.	Usuarios	17
2.3.3.	Uso del Internet y desigualdad digital en Ecuador	19
2.4.	Minería de datos.....	22
2.4.1.	Tipos de datos	22
2.4.2.	Tipos de modelos	23
2.4.3.	La minería de datos y el proceso de descubrimiento de conocimiento en bases de datos.....	24
2.4.3.1.	Fase de selección, limpieza y preparación de los datos	24
2.4.3.2.	Fase de minería de datos.....	25
2.4.4.	Herramienta SPSS	29
2.4.5.	Herramienta Weka.....	29
2.5.	Internet y rendimiento académico	29
2.6.	Proyectos relacionados	30
2.6.1.	Proyecto: El uso de internet para la interacción en el aprendizaje: un análisis de la eficacia y la igualdad en el sistema universitario Catalán. (Castaño Muñoz, 2011)	30
2.6.2.	Proyecto: Relación entre el uso de internet y el logro académico (Ayala, 2007)	30
2.6.3.	Proyecto: 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 Díaz, 2012).....	31
2.6.4.	Proyecto: Las técnicas de Clustering en la personalización de sistemas de e-learning (Farias, Durán, & Figueroa, 2008).	32
CAPITULO III: METODOLOGÍA		33
3.1.	Población y muestra	34
3.2.	Recolección, Procesamiento y Análisis de datos	35
3.2.1.	Recolección de Información.....	35
3.2.2.	Procesamiento y Análisis de Datos.....	38
3.2.3.	Agrupar a los estudiantes mediante análisis de Conglomerados	38
3.2.4.	Verificar el cluster mediante análisis Discriminante.....	39
3.2.5.	Elaboración de tablas de doble entrada, para aplicar el estadístico de Chi cuadrado. (X ²)	39
3.2.6.	Comprobación de Hipótesis mediante Regresión Logística.	39
3.3.	Metodología para el uso de la herramienta Weka.	40
3.3.1	Integración y recopilación	41
3.3.2	Selección, limpieza y transformación de los datos.....	41

3.3.3	Minería de Datos y extracción de patrones.....	44
3.3.3.1	Algoritmos para selección de atributos.....	44
3.3.3.1.1	DataSet para la hipótesis uno	45
3.3.3.1.2	DataSet para la hipótesis dos	46
3.3.3.1.3	DataSet para la hipótesis tres	47
3.3.3.1.4	DataSet para la hipótesis cuatro	47
3.3.3.2	Algoritmo de agrupación (K-means).....	48
3.3.4	Evaluación e Interpretación	51
CAPÍTULO IV: RESULTADOS		52
4.1	Descripción de la muestra	53
4.2	Conocimiento y preferencias del estudiante.....	55
4.3	Uso del internet en el ámbito académico y no académico.....	58
4.3.1	Usos académicos	58
4.3.2	Usos no académicos	62
4.4	Uso de redes sociales	64
4.5	Rendimiento Académico	65
4.6	Relaciones entre variables.....	65
4.7	Perfiles de estudiantes	66
4.7.1	Perfil de estudiantes en función a las actividades académicas	66
4.7.2	Perfil de estudiantes en función a las actividades de entretenimiento.....	68
4.7.3	Perfil de estudiantes en función al nivel, uso de dispositivos	70
4.8	Comprobación de las hipótesis.....	72
4.8.1	Perfil académico y perfil de entretenimiento	72
4.8.2	Usos de internet y rendimiento académico	76
4.9	Evaluación e Interpretación con el uso de la Herramienta Weka	78
4.9.1	Resultado Algoritmos para selección de atributos	78
4.9.2	Resultado del algoritmo SimpleKMeans.	78
CAPÍTULO V: DISCUSIÓN DE RESULTADOS.....		94
CONCLUSIONES		102
REFERENCIAS		¡Error! Marcador no definido.
ANEXOS.....		108

ÍNDICE DE TABLAS

Tabla 1: Descripción de las variables sociodemográficas	35
Tabla 2: Descripción de las variables de conocimiento y preferencias del estudiante en el uso de Internet.....	36
Tabla 3: Descripción de las variables referentes al uso de Internet Académico.	36
Tabla 4: Descripción de las variables referentes al uso de Internet en entretenimiento y diversión.	37
Tabla 5: Descripción de las variables; nivel de uso de dispositivos y materias que el estudiante se matriculó.....	38
Tabla 6: Variables consideradas para el análisis de Hipótesis 1 y 3	42
Tabla 7: Variables consideradas para el análisis de Hipótesis 2 y 4	43
Tabla 8: Variable a ser recodificada para medir los usos de Internet	43
Tabla 9: Umbrales para la construcción de la variable usos_internet_academico.....	44
Tabla 10: Resultados obtenidos de los algoritmos de selección de atributos para variables académicas	45
Tabla 11: DataSet para la comprobación de la Hipótesis uno, con Weka	46
Tabla 12: Resultados de selección de atributos para variables de entretenimiento.....	46
Tabla 13: DataSet para la comprobación de la Hipótesis dos, con Weka.....	47
Tabla 14: DataSet para la comprobación de la Hipótesis tres, con Weka	47
Tabla 15: DataSet para la comprobación de la Hipótesis cuatro, con Weka	48
Tabla 16: DataSet Configuración de opciones en el algoritmo SimpleKMeans.	48
Tabla 17: DataSet Resultados de experimentos para la hipótesis 1.....	49
Tabla 18: DataSet Resultados de experimentos para la hipótesis 2.....	49
Tabla 19: DataSet Resultados del segundo análisis de experimentos para la hipótesis 2....	50
Tabla 20: DataSet Resultados de experimentos para la hipótesis 3.....	50
Tabla 21: DataSet Resultados de experimentos para la hipótesis 4.....	51
Tabla 22: DataSet Resultados del segundo análisis de experimentos para la hipótesis 4....	51
Tabla 23: Cantidad de días a la semana que se conectan a Internet los estudiantes	56
Tabla 24: Distribución de estudiantes por el nivel de conocimiento en Internet.....	56
Tabla 25: Años de uso de internet del estudiante.	57
Tabla 26: Ingreso a la plataforma virtual	58
Tabla 27: Consultas al profesor	58
Tabla 28: Descargas de recursos educativos	59
Tabla 29: Uso del chat académicos por horas al mes	61
Tabla 30: Uso del chat por diversión por horas a la semana.....	62
Tabla 31: Uso de redes sociales por horas a la semana.....	63

Tabla 32: Uso de herramientas para descargar música, videos y programas	63
Tabla 33: Uso de la cuenta YouTube para entretenimiento	64
Tabla 34: Porcentajes de exactitud de la clasificación-Académica.....	67
Tabla 35: Porcentajes de exactitud de la clasificación-Entretenimiento	69
Tabla 36: Porcentajes de exactitud de la clasificación-Nivel uso de dispositivos.....	71
Tabla 37: Coeficientes del modelo de regresión de perfiles de uso de Internet en actividades académicas e Ingresos	74
Tabla 38: Coeficientes del modelo de regresión de perfiles de uso de Internet en actividades de entretenimiento e Ingresos.....	75
Tabla 39: Coeficientes del modelo de regresión de rendimiento académico	76
Tabla 40: Coeficientes del modelo de regresión de rendimiento académico e Uso Internet entretenimiento	77
Tabla 41: Resultados de mejor ajuste de Umbral.....	78

ÍNDICE DE FIGURAS

Figura 1: Perfil de usuario redes sociales	19
Figura 2: Tipos de acceso a internet en los hogares ecuatorianos.....	20
Figura 3: Porcentaje de personas que han utilizado internet por área en los últimos 12 meses.....	21
Figura 4: Porcentaje del uso de Internet por lugar de acceso	21
Figura 5: Proceso de Knowledge Discovery in databases.....	24
Figura 6: Proceso de determinación de Grupos k-means	27
Figura 7: Fases del proceso de extracción del conocimiento	41
Figura 8: Distribución de estudiantes por género.....	53
Figura 9: Distribución de estudiantes por titulación académica.....	54
Figura 10: Distribución de estudiantes por edad	54
Figura 11: Distribución de estudiantes por nivel de ingresos económicos.....	55
Figura 12: Distribución de estudiantes por lugar de conexión.....	55
Figura 13: Distribución de estudiantes por tiempo de conexión.....	57
Figura 14: Distribución de participación de los estudiantes en foros virtuales al mes.....	59
Figura 15: Videos académicos que el estudiante mira en YouTube al mes.....	60
Figura 16: Tweets académicos que realiza el estudiante al mes.....	60
Figura 17: Información académica que buscan los estudiantes en internet.....	61
Figura 18: Nivel de uso de acceso de los estudiantes a la biblioteca virtual.....	62
Figura 19: Amigos o seguidores de los estudiantes en las redes sociales	64
Figura 20: Distribución del rendimiento académico de los estudiantes	65
Figura 21: Variables influyentes en el uso de internet académico.....	67
Figura 22: Perfiles de estudiantes en función de las actividades académicas.....	68
Figura 23: Variables influyentes en el uso de internet de entretenimiento.....	69
Figura 24: Perfiles de estudiantes en función de las actividades de entretenimiento	70
Figura 25: Variables influyentes en el Uso de dispositivos.....	71
Figura 26: Perfiles de estudiantes en función al nivel de uso de dispositivos.....	72
Figura 27: Vista general de las variables que intervienen en la Hipótesis uno	79
Figura 28: Resultados de la ejecución del algoritmo- SimpleKMeans-Hipótesis uno.....	80
Figura 29: Perfil de estudiantes en función de los usos de Internet en actividades académicas	81
Figura 30: Vista general de las variables que intervienen en la Hipótesis dos-Experimento1	82
Figura 31: Resultados de la ejecución del algoritmo- SimpleKMeans-Hipótesis dos-experimento 1	83

Figura 32: Perfil de estudiantes en función de los usos de Internet en entretenimiento con Weka- Experimento 1	83
Figura 33: Vista general de las variables que intervienen en la Hipótesis dos-Experimento 2	84
Figura 34: Resultados de la ejecución del algoritmo- SimpleKMeans-Hipótesis dos-experimento 1	85
Figura 35: Perfil de estudiantes en función de los usos de Internet en entretenimiento con Weka- Experimento 2	86
Figura 36: Vista general de las variables que intervienen en la Hipótesis tres	87
Figura 37: Resultados de la ejecución del algoritmo- SimpleKMeans-Hipótesis tres.....	88
Figura 38: Uso académico de internet y rendimiento académico con Weka.....	89
Figura 39: Vista general de las variables que intervienen en la Hipótesis cuatro-Experimento1	89
Figura 40: Resultados de la ejecución del algoritmo- SimpleKMeans-Hipótesis cuatro-Experimento1	90
Figura 41: Uso de internet en entretenimiento y rendimiento académico-Experimento 2	91
Figura 42: Vista general de las variables que intervienen en la Hipótesis cuatro	91
Figura 43: Resultados de la ejecución del algoritmo- SimpleKMeans-Hipótesis cuatro-Experimento 2.....	92
Figura 44: Uso de internet en entretenimiento y rendimiento académico-Experimento 2	93

TERMINOLOGÍA

TIC.- Tecnologías de la información y la comunicación

INEC.- Instituto Nacional de Estadística y Censos

NRI.- Índice de disponibilidad de tecnología 'Networked Readiness Index' (NRI, que podría traducirse como Índice de Disponibilidad de Conectividad).

CONATEL.- Agencia de Regulación y Control de las Telecomunicaciones

KDD.- Descubrimiento de conocimiento en base de datos" (Knowledge Discovery in

El odd ratio (OR).- La razón de momios (RM) o razón de oportunidades -en inglés, odds ratio (OR), es una medida estadística utilizada en estudios epidemiológicos transversales y de casos y controles, así como en los metaanálisis.

ÍNDICE DE ANEXOS

Anexo 1	109
Anexo 2	111
Anexo 3	115
Anexo 4	116
Anexo 5	120
Anexo 6	121
Anexo 7	124
Anexo 8	125
Anexo 9	126
Anexo 10	127
Anexo 11	129
Anexo 12	131
Anexo 13	133
Anexo 14	135
Anexo 15	136
Anexo 16	141
Anexo 17	143

RESUMEN

En el presente trabajo de titulación, se realizó un estudio en donde se detallan algunas características de la brecha digital; específicamente se analizó la incidencia de la variable, nivel de ingresos económicos del estudiante sobre los usos de internet en actividades académicas y de entretenimiento, a su vez, con estos resultados obtenidos se determinó la relación que existe entre la variable uso de internet y rendimiento académico del estudiante. Para llegar a los objetivos planteados se utilizó técnicas de minería de datos en dos herramientas: en el primer software SPSS, se realizó análisis cluster para obtener perfiles de usuarios; estos perfiles se comprobaron con análisis discriminante y, finalmente, se aplicó análisis de regresión logística para comprobar las hipótesis; con la herramienta Weka, se procedió aplicar básicamente, algoritmos de selección de atributos para obtener las variables más relevantes, agrupamiento k-means para clasificar a los estudiantes y posteriormente, se interpretó los resultados.

PALABRAS CLAVES: Brecha digital, uso de Internet, información, rendimiento académico, minería de datos, incidencia.

ABSTRACT

In the present degree work, I did a study where some characteristics of the digital divide are detailed; specifically I analyzed the impact of the variable which is referred to the level of the students' economic incomes linked to the use of internet in academic and entertainment activities, at the same time, with the results gotten I analyzed the relationship between the variable: "use of Internet " and the student's academic performance. To reach the objectives proposed I used data mining techniques using two tools: the first SPSS software, a cluster analysis was performed to obtain user profiles; these profiles were tested with discriminant analysis and finally a logistic regression analysis was applied to test the hypotheses. On the other hand, I used the Weka tool to apply feature selection algorithms to get the most relevant variables, k-means clustering to classify the students and finally to interpret the results.

Keywords: digital divide, Internet use, information, academic performance, data mining, incidence.

INTRODUCCIÓN

Las tecnologías de la información y la comunicación (TIC), ocupan un lugar importante en la sociedad y economía de un país; y han venido evolucionando de manera muy rápida. A pesar de la presencia e influencia del uso de Internet en todos los ámbitos, existen grupos y actores sociales que no pueden dar el mayor aprovechamiento que éste ofrece, las causas son diversas y generalmente dependen de la estructura social y económica, esto genera lo que hoy conocemos como el fenómeno de brecha digital.

Esta investigación tiene como propósito comprender y analizar las diferencias del uso de Internet en la universidad Técnica Particular de Loja; las relaciones de la variable sociodemográfica ingresos económicos del estudiante, con el uso que le dan al Internet, en el ámbito académico y de entretenimiento; y a su vez la incidencia de éste sobre el rendimiento académico.

Con la ayuda de las técnicas de minería de datos, como análisis: clustering, discriminante, regresiones, se logró obtener información que permita determinar las relaciones existentes entre el uso de Internet y los ingresos económicos de los estudiantes, de forma similar se procedió a buscar relaciones entre los usos de Internet y el rendimiento académico.

En el presente proyecto de titulación está formado por cinco capítulos; en el Capítulo I muestra los objetivos, las preguntas de investigación e hipótesis que están relacionadas con las preguntas, así como, el procedimiento que se realizó.

El Capítulo II, describe el estado del arte del presente proyecto, destacando los aspectos, factores que determinan la brecha digital y el efecto del uso de tecnología sobre el rendimiento académico.

El Capítulo III, presenta la metodología utilizada para la realización de la investigación así como, el cálculo de la muestra, población, entrevista, y, finalmente la descripción de las técnicas de minería de datos empleadas: análisis cluster, análisis discriminante y regresión logística.

El Capítulo IV, se centra en mostrar los resultados obtenidos a través de las encuestas, así mismo, se relaciona en primer lugar las ingresos y los perfiles de los estudiantes y en segundo lugar, los perfiles de estudiantes y el rendimiento académico; además se muestra

la comprobación de hipótesis a través del modelo de regresión logística en el caso de Spps, y K-means en Weka.

Por último en el Capítulo V, se presentan la discusión de resultados del proyecto.

La información recogida para determinar la incidencia del uso de Internet con la variable ingresos económicos; y, rendimiento académico del estudiante, servirá como un elemento transcendental para la toma de decisiones al interior de la institución.

CAPITULO I: PLANTEAMIENTO DE LA INVESTIGACIÓN

1.1. Objetivos

- Determinar los usos de internet en estudiantes en la Universidad Técnica Particular de Loja
- Determinar el nivel de incidencia de la variable Ingresos sobre los usos de Internet. (aprendizaje y entretenimiento)
- Relacionar el rendimiento académico y los usos de internet. (aprendizaje y entretenimiento).

1.2. Preguntas de Investigación

- ¿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?
- ¿Cómo se relacionan el rendimiento académico y los usos de internet en actividades académicas y de entretenimiento?

1.3. Hipótesis

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

1.4. Procedimiento

Fase 1: Sistema de Información

Se elaboraron instrumentos de recogida de datos: encuesta y cuestionario en línea; y, se aplicaron a una muestra de 500 estudiantes, se utilizó métodos estadísticos y de minería de datos para la realización de categorizaciones y regresión logística para establecer modelos.

Fase 2: Preparación de los datos

Con los datos obtenidos de las encuestas aplicadas a los estudiantes, se procedió a digitalarlos en línea para luego ser tratados con las técnicas estadísticas y de minería de datos que nos permitió descubrir el reconociendo relevante y la calidad de los datos.

Fase 3: Minería de Datos e Interpretación y evaluación

Se procedió aplicar las técnicas estadísticas y de minería de datos, se estableció relaciones entre el nivel de ingresos económicos de los estudiantes y los usos de Internet, para determinar el rendimiento académico se lo realizó de forma similar; todo esto con la ayuda de análisis: cluster, discriminante, regresiones; y, tareas de minería de datos, es decir, el algoritmo K- means. Con los patrones obtenidos después de realizar varios experimentos se dio paso a documentar las relaciones y hallazgos.

Fase 4: Generación de Informe

Finalmente, se documentó el trabajo, el cual consta de los siguientes apartados: introducción, marco teórico, metodología, resultados, discusión de resultados, conclusiones y referencias.

CAPÍTULO II: MARCO TEÓRICO Y CONCEPTUAL

2.1. Importancia de las Tecnologías de Información y Comunicación.

Por medio del internet fluyen diversos tipos de información que se tornan relevantes para realizar investigaciones, continuamente los adelantos tecnológicos incrementan, a la par se crean herramientas para el manejo de la información, puesto que estas tecnologías están transformando la sociedad y modificando con mayor intensidad las formas de vida del ser humano.

El autor (Castaño Muñoz, 2010), plantea que existen otras dimensiones que se deben indagar: las habilidades en el uso de Internet, tiempo en la red y el tipo de uso que la gente hace de Internet. Para cada una de estas dimensiones es importante estudiar sus determinantes y sus consecuencias sociales.

El origen del internet ha desplegado el uso de las tecnologías de la información y la comunicación (TIC), en el ámbito de la enseñanza, generando una nueva acogida que la han convertido en una necesidad, por ello se establece a la par herramientas de trabajo básicas de interacción entre el educador y los alumnos.

En cambio, (Coll, 2011) expone que *“Las TIC están transformando escenarios educativos tradicionales que a su vez están haciendo aparecer otros nuevos”*, esto implica que las instituciones educativas formales tales como escuelas, institutos, universidades, etc., están evolucionando constantemente ante el creciente uso de la tecnología. La visión de nuevas necesidades se cubre con la aparición de nuevas tecnologías que conllevan a la creación de nuevos escenarios idóneos para la formación y el aprendizaje.

Como se menciona anteriormente, el uso de las TIC, constantemente han reflejado un gran impacto para fortalecer el proceso de enseñanza y aprendizaje en las universidades e instituciones educativas, cabe recalcar que así como tiene sus ventajas también tienen sus desventajas, ya que pueden ser utilizadas a favor o en contra, dependiendo de la cultura de cada persona y la calidad de la educación inculcada.

2.2. Brecha Digital

2.2.1. Introducción

Las TIC y sus características significativas representa la parte esencial en los cambios de la economía y de la sociedad a nivel mundial, pero las TIC no se han introducido de forma conjunta en la sociedad, es por ello que los sociólogos se centran en estudiar las diversas formas de desigualdad social que estarían asociadas a esta tecnología, estableciendo principalmente dos formas, la disponibilidad de medios económicos y medios tecnológicos de donde surge la denominada brecha digital.

El término brecha digital procede del inglés digital-divide, creado a mediados de los 90, cuyo uso oficial se dio a partir del año 1995 en el estudio germinal de la NTIA (National Telecommunications & Information Administration), donde se asumió el digital-divide como producto de analizar los reportes de acceso a internet; en los EE.UU. se encontró como factor común un menor acceso en los estratos sociales con bajos ingresos.

2.2.2. Definiciones

Con el tiempo el término de brecha digital ha ido evolucionando en busca de un mejor entendimiento, a continuación se presentan definiciones según varios autores.

Según, (Serrano & Martinez, 2003), expone que *“La brecha digital es la separación que existe entre las personas (comunidades, estados, países...) que utilizan las Tecnologías de Información y Comunicación (TIC), como una parte rutinaria de su vida diaria y considera aquellas que no tienen acceso a las mismas y que a pesar de disponer de ellas no saben cómo utilizarlas”*

La autora (Maggio, 2007), menciona que *“Es una expresión que hace referencia a las diferencias socioeconómicas entre aquellas comunidades que tienen accesibilidad a Internet y las que no, aunque tales desigualdades también se pueden referir a todas las nuevas tecnologías de información y la comunicación (TIC), como el computador personal, la telefonía móvil, la banda ancha y otros dispositivos.”*

En cambio, (Pippa Norris, 2001) considera que *“El concepto de brecha digital incluye tres diferentes aspectos: La división global, que se da entre países industrializados y los de bajo nivel de infraestructura. La división social enmarcada entre la diferencia de ricos y pobres en cada país. Y finalmente la división democrática que contempla la separación entre las personas que hacen uso de los recursos para involucrarse y participar en la vida pública”*

Otra definición de brecha digital es la que proponen los autores (DiMaggio, Hargittai, Celeste, & Shafer, 2004), quienes utilizan el término *digital inequality* o desigualdad digital para definir claramente toda aquella desigualdad social relacionada con la aparición de internet; definen 5 dimensiones que pueden generar desigualdad digital.

- ✓ Los medios técnicos o infraestructuras de conexión.
- ✓ La autonomía de uso.
- ✓ Los patrones de uso de internet.
- ✓ Las redes de soporte social.

- ✓ Las habilidades de uso.

Las diferentes formas de relación con la tecnología entre las personas que usan internet se denomina desigualdad digital (DiMaggio et al., 2004), estos autores quienes fueron los que plantearon una aproximación más exacta de brecha digital, resumieron que es un acceso característico de los usuarios en cuanto al uso que hacen de internet, tratando de pasar por las limitaciones de esta idea, y para quienes la brecha digital también incluye el buen uso de la tecnología y no solo el acceso a ella.

Este fenómeno de la brecha digital implica un deterioro en las oportunidades sociales, laborales y culturales de las personas, algunos autores identifican variables socioeconómicas. Para (Taylor Wallace, Dekkers, Marshall, & Zhu, 2003), existe una relación directa entre los niveles de ingreso de las familias y los niveles de uso del internet, lo que determina que la desigualdad digital sea una extensión de la desigualdad social y sus efectos vayan más allá de la dicotomía de estar o no conectado.

Considerando los puntos de vista de varios autores sobre el tema de brecha digital, se observa en general una diferencia digital en la sociedad, analizada principalmente desde dos ámbitos:

- Acceso, habilidades, principales usos e intensidad de uso.(Castaño Muñoz, 2010)
- Variables socioeconómicas, donde predominan los niveles de ingreso de las familias, educación, raza, género, ocupación, edad, estructura familiar, entre otras. (Dimaggio & Hargittai, 2001) & (Taylor Wallace et al., 2003) & (DiMaggio et al., 2004)

2.2.3. Aspectos de la Brecha Digital

En el año 2014, la Unión Internacional de Telecomunicaciones a través de su evento "*Building digital bridges*" analizó el concepto de brecha digital, considerando la diferencia en la posibilidad de conectarse y la divergencia en las velocidades de conexión. (UIT, 2014).

En este caso se establecen tres tipos de brecha digital; la de acceso, basada en la diferencia entre las personas que pueden acceder a las TIC y las que no; la de uso, basada en las personas que pueden y las que no saben utilizarlas; y las de calidad de uso, basada en las diferencias entre los mismos usuarios (Cho, 2004).

2.2.3.1. Brecha Digital de Acceso

Esta brecha considera, la dificultad que las personas tienen para disponer de una computadora con acceso a internet, incluye a su vez las redes avanzadas como tener servidores.

2.2.3.2. Brecha Digital de Uso

La brecha digital de uso, se basa en la capacidad o dificultad que tienen las personas para utilizar las TIC; al existir barreras para el acceso a estas tecnologías, y el conocimiento de cómo usarla, se ha implementado una alfabetización digital, con la finalidad de disminuir la brecha digital.

2.2.3.3. Brecha Digital de Calidad de uso

La brecha digital de calidad de uso, es la ausencia de posibilidades que tienen los internautas para acceder a diferentes herramientas que ofrece el internet, es decir, la limitación que impone la falta de conocimientos sobre la existencia de dichas herramientas para buscar información, por ejemplo, de medicina, estadísticas, redes sociales, o incluso para encontrar oportunidades de trabajo mediante la red.

La brecha digital se ha ido modificando con el pasar del tiempo, varios autores han empezado a introducir el desarrollo de las capacidades y habilidades requeridas para utilizar las TIC en educación y capacitación (ITU, 2005), de tal forma, se le incorpora al concepto de brecha digital tres principales enfoques detallados a continuación,

- a. **El enfoque hacia la infraestructura:** Se refiere a la posibilidad o dificultad de disponer de computadoras conectadas a la red mundial, debe entenderse a su vez la diferencia entre acceso formal y acceso real.

El acceso formal a internet se define como la disponibilidad de las infraestructuras necesarias para conectarse a la red. Para permitir este acceso, es necesario, poner al alcance de los individuos las conexiones a internet, superando así la barrera del acceso material. (van Dijk & Hacker, 2003).

Por otro lado, *el acceso real a internet*, va más allá del acceso formal, en este caso se considera que los internautas al disponer de internet no necesariamente lo utilizan, es decir, las TIC pueden estar presentes pero el usuario puede como no hacer uso de ellas.

- b. **El enfoque hacia la capacitación:** Representa las capacidades y potencialidades de usar las TIC. Al existir una gran diferencia asociada con las habilidades y capacidades de los internautas para utilizar las diferentes tecnologías, se empieza a difundir el concepto de alfabetización digital que va asociado con la brecha digital.

- c. **El enfoque hacia el uso de los recursos:** Considera la posibilidad que tienen los internautas, para utilizar los recursos disponibles en la red, estos se asocian con nuevas oportunidades para el desarrollo de la sociedad, puesto que las TIC son consideradas motores para mejorar la economía, ya sea para la atención médica en línea, o para disfrutar de nuevas formas de trabajo o distracción.

2.2.4 Factores que determinan la Brecha Digital.

Con base en la literatura existente se señalan diversos factores que interfieren en el aprendizaje, tales como las diferencias de acceso y uso a las tecnologías (DiMaggio et al., 2004).

La Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO, 2009), menciona que la exclusión de la información no sólo es una cuestión de acceso y conexión, sino también de contenidos.

Según datos publicados en (Cuneo, 2002), solamente el 11% de la población mundial tiene acceso a Internet. El 90% de ellos están en países desarrollados: 30% en América del Norte, 30% en Europa, 30% en Asia y el Pacífico y el 10% restante se encuentra entre el 82% de la población mundial.

En un artículo del 2011 (Proenza, Bastidas-Buch, & Guillermo, 2011), se concluye que el acceso a internet es sumamente desigual entre países desarrollados y subdesarrollados, por ejemplo, sólo el 2% de los habitantes de América Latina y el Caribe usan el internet comparado con el 40% en Estados Unidos y 36% en Canadá. En los países de Centroamérica tienen 7% de la población de América Latina y el Caribe, pero solo 1% de los dominios y una presencia imperceptible en la nueva economía.

El informe de la Unesco señala que la brecha digital se debe a múltiples causas, siendo principalmente el deficiente elemento estructural, dentro del cual se presentan los siguientes escenarios:

- 2000 millones de personas (1/3 de la población del mundo) no tienen acceso a electricidad.
- El coste de las telecomunicaciones es muy alto en el Tercer Mundo.
- Los ordenadores tiene un costo elevado y el suministro de los servicios asociados sólo llegan a las grandes ciudades, por lo que su extensión es baja en zonas rurales.
- La brecha digital como la brecha cognitiva tienen relación con los obstáculos educativos, culturales y lingüísticos que hacen del internet un objeto extraño e

inaccesible para las poblaciones que han quedado limitadas en los márgenes de la mundialización. (UNESCO, 2005).

A continuación se presentan los factores que influyen en la brecha digital.

- **Recursos económicos.-** Son un factor importante para que las empresas de un país adopten o no a las TIC dentro de sus funciones; Sin embargo, la tecnología puede ser una variable independiente como dependiente del crecimiento económico, el precio de un ordenador varia en los países de América del Sur, así como el costo elevado de las inversiones en infraestructuras, esto representan un poderoso factor de desigualdad.
- **La Geografía.-** La asimetría entre las ciudades y el campo crea situaciones de profunda desigualdad. Los países desarrollados tienen mayor acceso que los subdesarrollados con infraestructuras de baja calidad y menor capacidad
- **La edad.-** Los jóvenes tienen un uso elevado de las innovaciones tecnológicas y sus aplicaciones, pero constituyen un público vulnerable ante dificultades económicas y sociales. A su vez, el lento aprendizaje de las personas de cierta edad, ante el avance de las tecnologías se torna un obstáculo insuperable, por la carencia de estructuras de formación adecuadas.
- **El sexo.-** Las desigualdades entre hombres y mujeres en el ámbito de las nuevas tecnologías es otra faceta de la brecha digital, casi los dos tercios de los analfabetos del mundo son mujeres según la (UNESCO, 2009). En los países en desarrollo, una mujer de cada dos, por término medio no sabe leer; aunque, en los industrializados un número importante de mujeres son usuarias de Internet, en estos también se evidencian una serie de desventajas que les impide acceder a las nuevas tecnologías.
- **La lengua.-** Representa un obstáculo importante para la participación de todos en las sociedades del conocimiento. El auge del inglés como vector de la mundialización restringe la utilización de los demás idiomas en el ciberespacio.
- **La educación y la procedencia sociológica o cultural.-** La evolución de la sociedad "postindustrial" necesitará inversiones considerables en educación y formación. Puesto que se evidencia que existe una íntima conexión entre la sociedad de la información y las sociedades del conocimiento
- **El empleo.-** En muchos países, el acceso a internet sólo se efectúa en los lugares de trabajo y los "cibercafés", que distan mucho de estar al alcance de todos los bolsillos. La brecha tecnológica, a menudo va unida a la que resulta en materia de empleo.

- **La integridad física.**-Los discapacitados acumulan desventajas económicas, culturales o psicológicas que contribuyen a ahondar la brecha digital. Además, las discapacidades físicas en sí mismas representan un importante obstáculo para la utilización de los ordenadores. No obstante, es preciso reconocer los esfuerzos de los constructores para crear instrumentos que facilitan la utilización de los ordenadores por los discapacitados, por ejemplo la posibilidad de acceso a menús contextuales utilizando el teclado con una sola mano.

2.3. Uso de Internet

El Internet se ha convertido en una importante herramienta de acceso a la información. La red de internet contribuye al enriquecimiento de conocimientos y educación de toda la nación ayudando a los individuos a mejorar su nivel de vida.

A su vez es considerado una matriz global de redes de computadoras interconectadas entre sí, que utiliza el protocolo de internet (IP) para comunicarse una con la otra. IP forma parte del protocolo TCP/IP (Transfer Control Protocol/Internet Protocol) el cual se desarrolló para interconectar equipos de cómputo de diversas arquitecturas. (Serrano & Martinez, 2003)

El uso de internet, se ha ido consolidando a través del tiempo, su aceleración es extraordinaria convirtiéndose en una nueva plaza para el desarrollo de las naciones, que establece nuevas formas de relación social, ya que es un medio de comunicación masivo que permite intercambiar información de diversas formas.

Los buscadores contribuyeron a la percepción generalizada de Internet, como un inmenso depósito de información en el que se torna posible encontrar lo que buscamos. Hoy internet es mucho más que la pantalla de acceso a un buzón de correo electrónico o a un buscador. La red actualmente es un espacio de intercambio, de: creación, conocimientos compartidos, relación social, consecución de objetivos personales, económicos, sociales, educativos, etc. (Duart, 2010)

2.3.1. Uso y gratificación

El uso del internet se ha convertido en un medio de comunicación esencial entre las personas, en vista que se puede intercambiar todo tipo de información. Es el internauta el responsable de buscar o encontrar el contenido que requiere, es por esto que la teoría de usos y gratificaciones ya ha empezado abordarse dentro del tema de internet, ya que esta teoría nos ayuda a comprender las funciones del medio para la persona y la sociedad.

La teoría de usos y gratificaciones parte del supuesto de los usuarios como ente activo y racional que buscan contenidos específicos de medios (Blumler & Katz, 1974), todo esto con el fin de satisfacer las necesidades y las motivaciones del consumidor, es decir, explica el comportamiento entre individuos respecto al consumo de medios.

Las ideas básicas de este paradigma de usos y gratificaciones, se basan en cinco supuestos desarrollados por (Katz, Blumer, & Gurevitch, 1973).

1. El público es activo y se acerca a los medios para lograr ciertos objetivos. Es decir, que los usuarios usan los medios para obtener así beneficios. Los internautas son los que deciden a que sitio web acceden.
2. Cada miembro de la audiencia tiene la iniciativa de vincular la gratificación de una necesidad con la elección de un medio, la audiencia es quien usa el medio a su beneficio y tiene el control de lo que quiere ver. Internet es un medio disponible para satisfacer las necesidades del público, pero entrega explícitamente el control de lo que muestra a sus usuarios.(Cristián Ayala, 2007).
3. Hay una competencia entre los medios y otras fuentes a la hora de procurar unas gratificaciones a la gente. Cada usuario tiene diferente necesidad y cada medio puede satisfacer solo algunas.
4. Muchos de los objetivos del uso de los medios masivos, pueden derivarse de datos aportados por los mismos integrantes individuales del público. Los datos del público sobre las gratificaciones de los medios suscita problemas de validez; una limitación en el estudio de usos y gratificaciones es el hecho de haber trabajado firmemente con el mundo más directo de lo manifiesto.
5. Los juicios de valor, acerca de la calidad cultural e influencia nociva o no de los contenidos de los medios resultan inválidos y presuntuosos, si se renunciara a conocer lo que declaran al respecto los miembros del público (Dader, 1992). El valor está en la elección de cada usuario, al elegir un determinado medio.(Cristián Ayala, 2007).

Con base en la literatura antes expuesta se puede deducir que los medios se acomodan al gusto del consumidor, satisfaciendo así, las necesidades existentes. Actualmente hay dos preguntas que esta teoría trata de responder: *¿Por qué la gente usa los medios?* y *¿Cómo la gente construye sentido con los medios?*

Se pretende encontrar el camino de los procesos cognitivos, que se encuentran en el uso de los medios.

Investigadores como (LaRose, Mastro, & Eastin, 2001; Sorice, 2005), han emprendido el estudio de los usos y gratificaciones para entender el uso de internet a través de una teoría cognitiva social, con el objetivo de reducir incertidumbres que surgen al homogenizar la audiencia de internet y para explicar el uso del medio en términos de las gratificaciones que buscan los usuarios con su uso; por lo tanto la teoría social-cognitiva considera la autoayuda junto con el apoyo social adecuado.

También (LaRose et al., 2001), creó medidas para auto-eficacia y el auto-desprecio en donde relacionó la teoría de usos y gratificaciones con los resultados negativos que se producen en el comportamiento de adicciones en línea, y destaca el desarrollo de la auto-conciencia del uso del internet.

2.3.2. Usuarios

Para la (Unión Internacional de Telecomunicaciones, n.d.), “*Los usuarios de internet son personas con acceso a la red mundial.*” Mientras que según (Godoy & Herrera, 2003), el “*usuario es toda persona que haya utilizado internet en los últimos 90 días en cualquier lugar y cualquier cantidad de tiempo, aunque no tenga un computador conectado en su hogar.*”

Los perfiles de usuarios de internet varían de acuerdo a la sociedad que se encuentre, cada grupo de edad busca información de acuerdo a su interés. Las primeras generaciones que empezaron a nacer en un entorno digital, fueron las de 1978 y 1994, adoptando estas nuevas tecnologías.

En el artículo *Qué ocurre cuando se usa (y no se usa) internet: resultados del World Internet*, los autores (Godoy & Herrera, 2003), señalan que hay una gran variedad de estudios sobre el impacto de internet; pero muchos llegan a conclusiones contradictorias, sin embargo, existen factores que afectan las posibilidades de acceder a internet, aquí se detecta las dimensiones críticas de la tercera brecha digital tal como lo menciona (Dimaggio & Hargittai, 2001). Así mismo menciona los factores que afectan a los usos y gratificaciones que el usuario espera del internet.

- **Conocimiento de la tecnología.** La ignorancia tecnológica constituye el primer peldaño de la llamada brecha digital, que margina a porcentajes importantes de la humanidad de los beneficios de internet y otras tecnologías digitales y divide a la humanidad entre los que tienen (haves) y los que no tienen (have nots).
- **Habilidades computacionales.-** Estudios recientes la destacan como el principal factor que está frenando la adopción de internet, y que determina los primeros síntomas de estancamiento del crecimiento de la red en varios países. Este factor

está relacionado con la educación y el ingreso de las personas, por ende bastante ligado a las preocupaciones tradicionales sobre la brecha digital.

- **Acceso.-** Para la mayoría de los autores, estar conectado a internet (aunque no sea en el hogar) es lo mínimo para participar de la sociedad de la información.
- **Experiencia.-** Superado el problema del acceso, las personas modifican su uso en función de los años de experiencia en red. Ello no impide que algunas actividades, notablemente el correo electrónico, mantengan su importancia.
- **Factores internos a los usuarios.-** Ser hombre, participar en grupos sociales, estar empleado y una actitud positiva hacia la tecnología favorecen el uso. Tener una red de soporte social, pedir ayuda a un familiar, amigo o compañero de trabajo, mejora la disposición a la experimentación con TIC.
- **Factores ambientales externos.-** Las políticas públicas, las estrategias del sector privado, el entorno económico y la cultura del país afectan también la evolución de la tecnología. El estado de la economía o la cultura de un lugar generan un contexto capaz de alentar o deprimir el interés y los incentivos ante las TIC.

En Ecuador, de acuerdo con el (INEC, 2011), el perfil del internauta se define de la siguiente manera:

- Los hombres son quienes más utilizan la red para comunicarse, educarse, informarse y trabajar.
- Más hombres que mujeres se conectan desde su hogar.
- La mayoría de usuarios son jóvenes entre 16 y 24 años.

En el año 2014, Ecuador mejoró el índice de disponibilidad de tecnología (NRI) en red. Dentro de las telecomunicaciones, actualmente ocupa el puesto 82 entre 144 países estudiados; además se observa que el 45,1% de usuarios se conecta al internet y a las redes desde sus hogares. Así también, el 36,2% de las personas que utilizaron internet en 2012 fueron hombres, el 64,9% de internautas tienen entre 16 y los 24 años de edad, el 24,7% de hogares tiene computadoras de escritorio y el 9,8, portátiles (El Comercio). Según el INEC, en el año 2012 el internet es utilizado en mayor parte para acceder a las redes sociales, tal como lo indica la Figura 1.

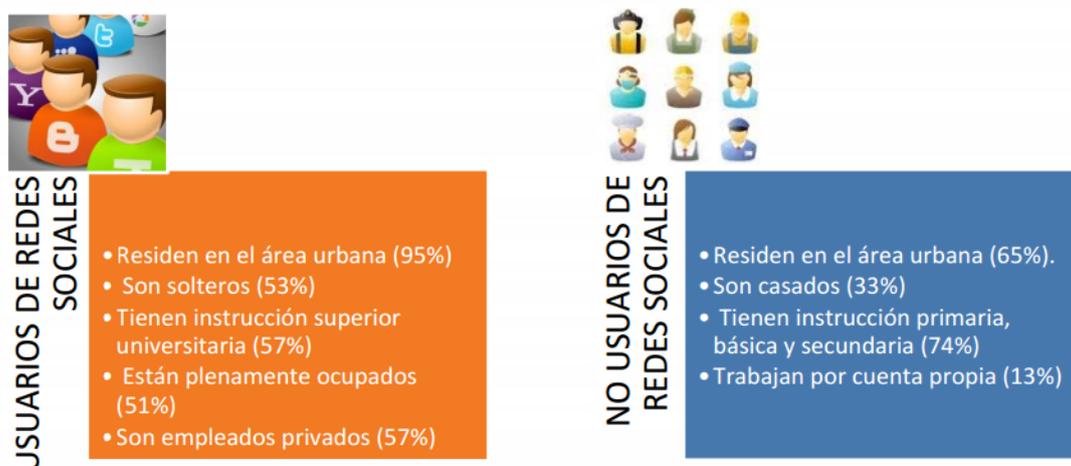


Figura 1: Perfil de usuario redes sociales
Fuente:(INEC, 2012)

2.3.3. Uso del Internet y desigualdad digital en Ecuador

El Consejo Nacional de Telecomunicaciones (CONATEL), es el encargado de regular el internet en nuestro país, esta entidad aprobó en el 2013 las reformas al reglamento de prestación del servicio de valor agregado de acceso al internet. (CONATEL, 2013b)

El beneficio directo para el usuario se refiere a una mayor oferta de prestación de acceso a internet, con mejores costos y ampliación del área de cobertura, expandiendo el servicio a zonas antes desatendidas y potenciando que todos los ciudadanos, independientemente de su ubicación, se incluyan en la sociedad del conocimiento, factores que disminuyen la brecha digital. (CONATEL, 2013a).

Según el marco regulatorio, en Ecuador existen dos tipos de acceso a internet:

- Conmutado (Dial-up)
- No Conmutado (Dedicado)

Usuarios Conmutados.- Son aquellos que acceden al servicio a través de líneas telefónicas conmutadas.

Usuarios Dedicados.- Son aquellos que acceden a la red a través de diferentes medios como cable, microondas, satélite o fibra óptica

La Figura 2, nos muestra el porcentaje de hogares que tuvieron acceso a internet mediante acceso conmutado y no conmutado durante los años 2010 al 2013, con base en la encuesta nacional de empleo, desempleo y subempleo-ENEMDU (INEC, 2013), se observó que el 28,3% de los hogares ecuatorianos tiene acceso a internet, de los cuales el 43,7% accede a

través de líneas telefónicas, porcentaje que ha disminuido en comparación al año 2012 que fue de 9,8% puntos, mientras que el 56,3% lo realizó mediante cable, banda ancha e inalámbrico, porcentaje que sufrió un incremento de 11.7% para el año 2013 (INEC, 2013).

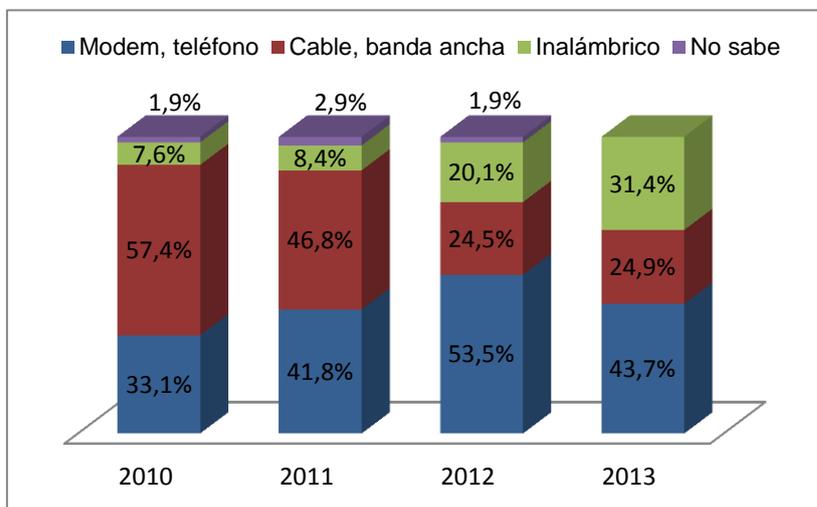


Figura 2: Tipos de acceso a internet en los hogares ecuatorianos.
Fuente:(INEC, 2013b)

Uno de los principales problemas para el acceso a internet en Ecuador son los aspectos regulatorios y los costos. Dentro de los aspectos regulatorios se puede mencionar la voluntad política, manejo de estado, tarifa plana, cargas tributarias y que el internet no es realmente un servicio público. En cuanto a costos, los principales aspectos son los valores de ISP, precios de última milla y cuantías de capacidad internacional (SUPERTEL, 2010).

Para el Instituto Nacional de Estadísticas y Censos (INEC), existe el 35% de acceso a internet en el Ecuador, esto representa que existen 5,2 millones de usuarios; en el 2013 se obtuvo que el 40,4% de la población de Ecuador ha utilizado internet en los últimos 12 meses, siendo el 47,6% de la población de la zona urbana que ha utilizado internet, frente al 25,3% del área rural, observándose una desigualdad digital. La tendencia del acceso a internet es hacia el alza, puesto que durante los diez últimos años el acceso a internet ha pasado de un 3% al 40%. Tal como lo indica la Figura 3.

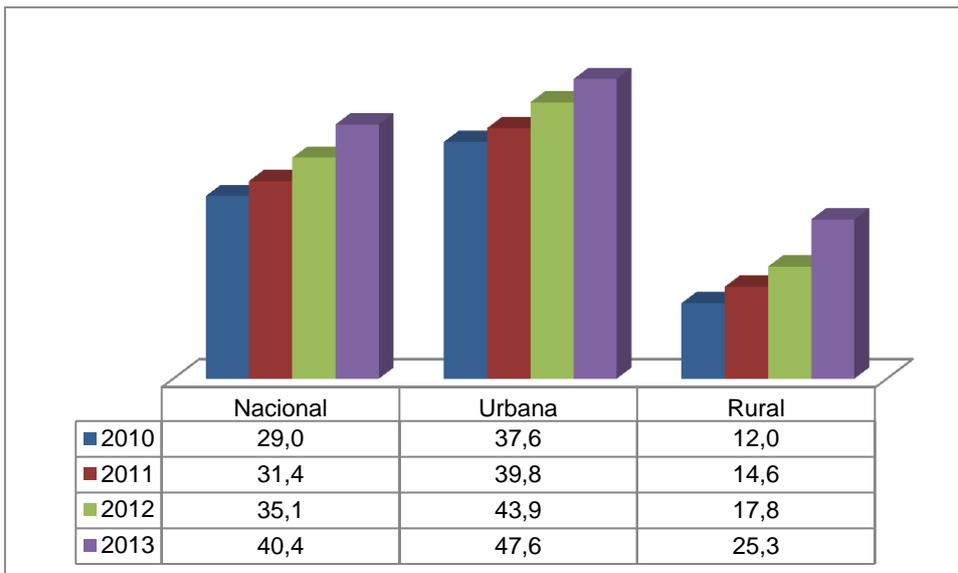


Figura 3: Porcentaje de personas que han utilizado internet por área en los últimos 12 meses.
Fuente:(INEC, 2013b)

Un dato muy importante que debe considerarse es el lugar de acceso a internet, se observa que el 45,1% lo hace desde su hogar, el 29,8% desde centros de acceso público, el 12,2% desde instituciones educativas, el 9,8% desde su lugar de trabajo y el 2,1% desde otros lugares. Ver Figura 4.

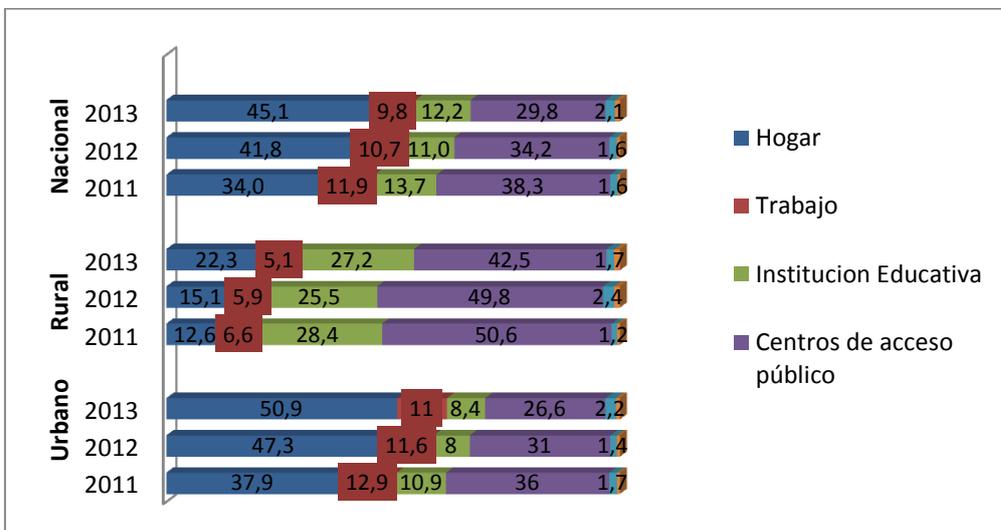


Figura 4: Porcentaje del uso de Internet por lugar de acceso
Fuente:(INEC, 2013b)

De acuerdo a todas estas estadísticas se puede observar que en Ecuador aún existe desigualdad digital, reflejándose a su vez la existencia de un analfabetismo digital, ya que alcanza el 62,9%, es decir, de cada 10 personas 6 habitantes no tiene conocimientos de las TIC.

2.4. Minería de datos

Actualmente la mayoría de empresas y organizaciones cuentan con sistemas de información que almacenan y generan grandes cantidades de datos, los cuales necesitan ser procesados para obtener información de utilidad y así tomar decisiones, para mejorar el negocio o servicio. A decir de (Hernández, Ramírez, & Ferri, 2007), los datos pasaron de ser el resultado histórico de los sistemas de información (producto) a ser una materia prima que hay que procesar para obtener un producto elaborado que se llama conocimiento, el cual permite tomar decisiones que permitan mejorar la productividad en el ámbito de donde se extrajo la información.

Formalmente (Clark & Robin, 2000), define a la minería de datos “como el proceso de extraer conocimiento útil y comprensible, previamente desconocida, desde grandes cantidades de datos almacenados en distintos formatos”.

En si la tarea principal de todo el proceso de minería de datos es descubrir modelos inteligibles, a partir de los datos se pueda explicar un comportamiento y desde los cuales se pueda tomar decisiones confiables que al final reporten un beneficio para la organización.

En definitiva la minería de datos nació como respuesta a las nuevas características de los datos (volumen y tipología) que la estadística tradicional no podía manejar.

2.4.1. Tipos de datos

La minería de datos puede aplicarse a cualquier tipo de dato, la única diferencia está en las técnicas de minería, las cuales varían dependiendo el tipo de dato. Con respecto al detalle de tipos de datos (Hernández et al., 2007), reconoce tres tipos de datos importantes:

2.4.1.1 Bases de datos relacionales: Son sistemas enfocados en almacenar de forma ordenada un conjunto de datos y sus respectivas relaciones. En el mercado se encuentran gran variedad de bases de datos como por ejemplo: Oracle, Mysql, Postgres, etc.

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

Una de las principales características de las bases de datos relacionales es la existencia de un esquema asociado, es decir, los datos deben seguir una estructura y son, por tanto, estructurados.

2.4.1.2. Otros Tipos de base de datos: Las bases de datos relacionales a pesar de ser las más utilizadas hoy en día, no son las únicas que existen, de hecho organizaciones con modelos de negocio particulares, necesitan de otros tipos de bases de datos que contienen datos complejos como:

- Bases de datos espaciales: Contienen información de datos geográficos, de ubicación, redes de transporte, información geográfica, etc. En si tienen información relacionada con el espacio en un sentido amplio.
- Bases de datos temporales: Contienen datos que incluyen muchos atributos relacionados con el detalle del tiempo: instantes, intervalos temporales.
- Bases de datos documentales: Contienen descripciones para los objetos; documentos estructurados (biblioteca digital de obras literarias), semi-estructurados (resúmenes) o estructurados (bases de datos de fichas bibliográficas)
- Bases de datos multimedia: Almacenan principalmente: imágenes, videos y audio, se caracterizan por tener gran capacidad de almacenamiento.

2.4.1.3. World Wide Web.- Este sin duda es el mayor repositorio de información, contiene diversos tipos de datos, aunque una desventaja es que la mayoría de información no está estructurada o en los mejores casos es semi-estructurada. Debido a la complejidad de su tipología de datos cuentan con su propia metodología de minería llamada Minería Web la cual se divide en tres categorías:

- Minería del contenido: para encontrar patrones de los datos de las páginas web.
- Minería estructurada: entendiendo por estructura los hipervínculos y URL's.
- Minería del uso, enfocada a descubrir información; que hace el usuario de las páginas web.

2.4.2. Tipos de modelos

En minería de datos se consideran como modelos a la forma en que se adquiere el conocimiento del proceso de minería, el cual puede ser en forma de relaciones, patrones o reglas inferidas de los datos previamente desconocidos. En si los modelos pueden ser de dos tipos:

- Modelos Predictivos: Sirven para estimar valores futuros o desconocidos, estos valores se consideran variables objetivo o dependientes, las cuales se obtienen a partir de otras variables o campos de las bases de datos, a éstas se consideran variables independientes o predictivas

- Modelos Descriptivos: Sirve para explorar las propiedades de los datos examinados, en si tratan de explicar el comportamiento de cierto fenómeno o comportamiento.

2.4.3. La minería de datos y el proceso de descubrimiento de conocimiento en bases de datos.

En (Hernández et al., 2007) se explica que el proceso de minería de datos no es más que un paso esencial de un proceso más amplio, cuyo objetivo principal es el descubrimiento de conocimiento en bases de datos. Es así, que últimamente se está utilizando y relacionando el término extracción o “descubrimiento de conocimiento en base de datos”(Knowledge Discovery in databases, KDD).

El KDD, se define como el proceso no trivial de identificar patrones válidos, novedosos, potencialmente útiles, y, en última instancia comprensibles a partir de los datos. Este proceso consta de un conjunto de pasos que se ejecutan de manera iterativa, porque a la salida de una etapa o fase puede volverse a fase anteriores porque en muchas de las ocasiones se necesita de varias iteraciones para extraer conocimiento relevante.

El proceso también se considera interactivo, porque necesita del criterio de un experto en el dominio del problema en las fases de pre-procesamiento de datos y validación del conocimiento extraído. Ver Figura 5

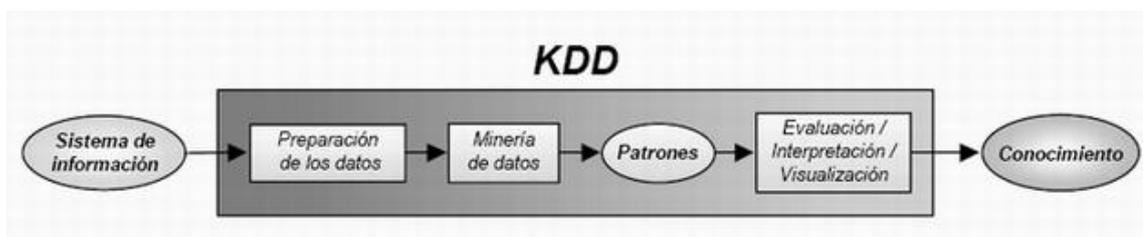


Figura 5: Proceso de Knowledge Discovery in databases
Fuente:(Hernández et al., 2007)

2.4.3.1. Fase de selección, limpieza y preparación de los datos

Descubrir conocimiento relevante y de calidad, no solo depende de la técnica de reconocimiento de patrones, sino también de la calidad y limpieza de los datos con que se trabaja.

Por tal razón, esta fase se hace necesaria e imprescindible, ya que se enfoca en seleccionar y preparar los datos o un subconjunto relevante de datos. En sí, los datos pueden tener varios problemas como irrelevancia, pueden tener información poco útil para el dominio del problema.

Otro problema son los datos que no se ajustan al comportamiento general, es decir, son excepciones de datos, los cuales no aportan información relevante, más bien pueden ocasionar errores o simplemente justifican un comportamiento diferente a los demás.

Finalmente están los datos perdidos, los cuales por algún error en la aplicación o dispositivo que los recolecta, no se puede contar con ellos. En este caso, lo que se acostumbra realizar es una proyección con los datos que se tiene a la mano. En conclusión, esta tarea cumple una función crucial, ya que de la calidad de los datos obtenidos en esta fase, se pueden obtener conocimiento relevante al final del proceso.

2.4.3.2. Fase de minería de datos

La fase de la minería de datos es la más representativa del KDD, se realiza el descubrimiento de conocimiento propiamente, el cual se obtiene de la construcción de un modelo basado en los datos recopilados. En definitiva el modelo es una descripción de los patrones y relaciones entre los datos que se pueden utilizar en predicciones para entender mejor los datos.

- **Tareas de minería de datos**

Se denomina tareas de minería de datos a los algoritmos necesarios para resolver un problema, estos algoritmos no pueden ser utilizados a todos los problemas ya que su aplicabilidad depende del tipo del problema.

Cada tarea pertenece a un modelo (predictivo o descriptivo) que revisamos en apartados anteriores y cuyo detalle se presenta a continuación:

- Tareas Predictivas:

Clasificación: A cada instancia se clasifica en una clase determinada, la cual se indica mediante el valor de un atributo que llamamos clase o instancia. El resto de los atributos se utiliza para predecir la clase.

- Tareas Descriptivas

Agrupamiento: En esta tarea se analizan los datos para generar grupos naturales. El proceso para llegar a formar los grupos se realiza basándose en el principio de maximizar la similitud entre los elementos de un grupo minimizando la similitud entre los distintos grupos.

Regresión: consisten en aprender una función real que asigna a cada instancia un valor real, el valor a predecir siempre tendrá que ser numérico.

Reglas de asociación: tiene como objetivo identificar tareas explícitas entre atributos categóricos. Las reglas de asociación no implican una relación de causa y efecto, puede no existir una causa para que los datos estén asociados

- **Reconocimiento de patrones**

A continuación, se describen algunas técnicas que resultan de gran utilidad para para realizar las tareas antes mencionada. En esta sección hablaremos de uno de los modelos de agrupamiento más utilizado.

Clustering.-Conocido también como agrupamiento, esta es una de las tareas más comunes en minería de datos. Se trata de encontrar grupos entre un conjunto de individuos. La distancia puede jugar en papel crucial, ya que los individuos similares deberían ir al mismo grupo.

Algoritmos de Partición.- Dado un conjunto D de n objetos en un espacio de d dimensiones, y tomando como parámetro de entrada el valor k , un algoritmo de particionamiento organiza los objetos dentro de k cluster, tal que sea minimizada la desviación total de cada objeto desde el centro de su cluster o desde una distribución de clusteres. La desviación de un punto puede ser computada en forma diferente según el algoritmo, y es llamada generalmente función de similitud.

Ejemplos: k-means, k-medoids, EM (Expectation Maximization)

K-medias: El algoritmo de las K-media desarrolladas por MacQueen en 1967 es uno de los algoritmos de aprendizaje no supervisado más simples y utilizados para resolver el problema de la clusterización. El procedimiento aproxima por etapas sucesivas un cierto número (prefijado) de clusteres haciendo uso de los centroides de los puntos que deben representar.

El algoritmo se compone de los siguientes pasos:

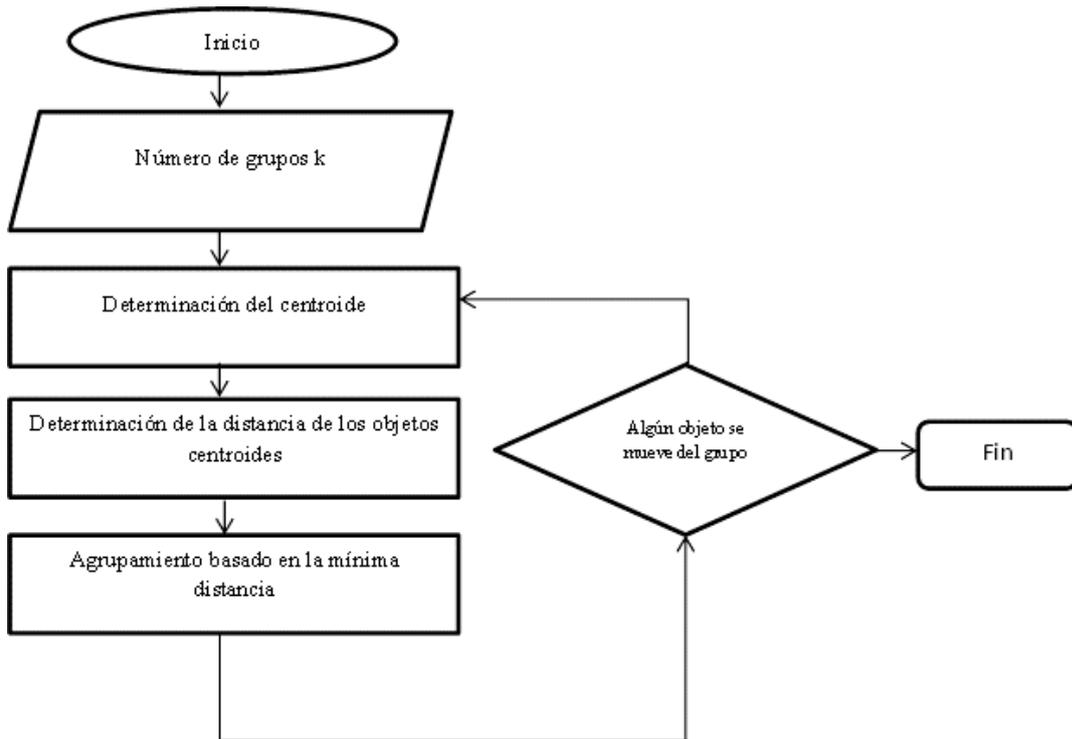


Figura 6: Proceso de determinación de Grupos k-means
Fuente:(Teknomos, 2007)

Algoritmos Jerárquicos (Hierarchical Algorithms): Un método jerárquico crea una descomposición jerárquica de un conjunto de datos, formando un dendograma, un árbol que divide la base de datos recursivamente en conjuntos cada vez más pequeños. El árbol puede ser formado de dos formas: de abajo hacia arriba (“bottom-up”) o de arriba hacia abajo (“top-down”).

En el caso “botton-up”, también llamado aglomerativo, se comienza con cada objeto formando un grupo por separado. Los objetos o grupos se combinan sucesivamente según determinadas medidas, hasta que todos los grupos se hayan unido en uno solo, o hasta que se cumpla alguna condición de terminación.

En el caso “top-down”, también llamado divisivo, se comienza con todos los objetos en el mismo cluster, y a medida que se va iterando, se dividen los grupos en subconjuntos más pequeños, según determinadas medidas, hasta que cada objeto esté en un cluster individual o hasta que se cumplan las condiciones de terminación.

Ejemplos: AGNES (Aglomerative NESTing), DIANA (Divisia ANALysis), CURE (Clustering

Using Representatives), CHAMELEON, BIRCH (Balanced Iterative Reducing and Clustering using Hierarchical)

Algoritmos Basados en Densidad (Density-Based Algorithms): La mayoría de los métodos de particionamiento, realizan el proceso de clustering en base a la distancia entre dos objetos. Estos métodos pueden encontrar sólo clusteres esféricos, y se les dificulta hallar clusteres de diversas Figuras. Otros algoritmos de Clustering han sido desarrollados con base en la noción de “densidad”. Estos generalmente estiman clusteres como regiones con gran densidad de objetos, separados de regiones de baja densidad de objetos (estos elementos aislados representan ruido). Este tipo de métodos es muy útil para filtrar ruido y encontrar clusteres de diversas formas. Ejemplos: DBSCAN (Density-Based Spatial Clustering of Applications with Noise), OPTICS(Ordering Points to Identify the Clustering Structure), DENCLUE (DENsity-based Clustering)

Algoritmos Basados en Grillas (Grid-Based Algorithms): Los métodos basados en densidad suelen tener grandes problemas cuando se trabaja con bases de datos muy grandes. Para mejorar la efectividad del Clustering, un método basado en grillas usa una estructura de grilla de datos. El método divide el espacio en un número finito de celdas, formando una grilla, en donde se realizan todas las operaciones del Clustering.

La mayor ventaja de este método es su veloz procesamiento del tiempo, el cual generalmente es independiente de la cantidad de objetos a procesar. Ejemplos: STING (A STatistical INformation Grid), WaveCluster, CLIQUE, etc.

Análisis cluster (clusterización): Comprobar si los modelos encontrados por los algoritmos de Clustering son correctos es una tarea difícil, ya que no existe una clase definida con la que se pueda comparar.

Sin embargo existen algunos criterios que nos permiten determinar aproximadamente la efectividad del modelo encontrado, los cuales pasamos a revisar a continuación:

- ✓ **Likelihood:** Es una medida que indican la verosimilitud que existe entre los datos. Matemáticamente se explica mediante un probabilidad estimada sea $p(x)$ la probabilidad de observar un punto en la posición x utilizando el modelo a evaluar, siendo p una función de densidad de probabilidad. Si el modelo es bueno se deben esperar probabilidades altas para los puntos que son observados.
- ✓ **Test de aplicación:** otra forma de evaluar si un modelo de agrupamiento es correcto es dividiendo la evaluación en dos grupos. En el primero se procesa todos los datos conocidos a manera de entrenamiento y luego se procesa los datos de tipo de evaluación si los resultados obtenidos entre la fase de entrenamiento y pruebas son similares o guardan relación, el modelo es correcto.

2.4.4. Herramienta SPSS

La herramienta que nos ayudará a realizar todo el trabajo de extracción de información será SPSS, el cual pasamos a revisar brevemente a continuación

Es un software global para el análisis de datos. SPSS Statistics puede adquirir datos de casi cualquier tipo de archivo y utilizarlos para generar informes tabulares, gráficos y diagramas de distribuciones y tendencias, estadísticos descriptivos y análisis estadísticos complejos.

2.4.5. Herramienta Weka

Weka es una completa suite para la extracción de conocimiento, desarrollado en el lenguaje Java y distribuido en forma Open Source en internet. El nombre del programa es la abreviatura de: Waikato Environment for Knowledge Analysis, y fue desarrollado por la universidad de Waikato de Nueva Zelanda.

Este programa tiene predefinido números algoritmos de análisis y aprendizaje automático, además de poderosas herramientas para los procesos previos a la minería de datos como son: Obtención y procesamiento de datos.

2.5. Internet y rendimiento académico

Una de las mayores preocupaciones existentes es el intento de conocer de qué manera puede contribuir la utilización de las TIC, en el mejoramiento del rendimiento académico de los alumnos.

Los autores (Jiménez, Izquierdo, & Blanco, 2000), se refieren al rendimiento académico como un constructor en el que no sólo se contemplan las aptitudes y la motivación del alumno; sino también otras variables intervinientes como los aspectos docentes, la relación profesor-alumno, el entorno familiar, etc.

En cambio (Forteza Méndez, 1975), conceptualiza el rendimiento académico como *“La productividad del sujeto, el producto final de la aplicación de su esfuerzo, matizado por sus actividades, rasgos y la percepción mas o menos correcta de los cometidos asignados”*., ante esta perspectiva el rendimiento académico se lo podría distinguir en dos aspectos: la medición cuantitativa (tales como notas) en las que el estudiante tiene que rendir pruebas que al final dependerá de que pase o no una materia; y de las variables cognitivas, afectivas como habilidades de pensamiento y conductas instrumentales para alcanzar las metas propuestas y elementos como la auto valoración, auto concepto, etc. (Alcalay & Antonijevic, 1987)

2.6. Proyectos relacionados

A continuación se muestran algunos proyectos relacionados con el uso del internet y el rendimiento académico. Estudios como este tipo no son nuevos, al contrario existen numerosos trabajos en la literatura que tratan de obtener conocimiento de cómo se ve afectado el rendimiento académico del estudiante, con base en las costumbres de uso de internet.

2.6.1. Proyecto: El uso de internet para la interacción en el aprendizaje: un análisis de la eficacia y la igualdad en el sistema universitario Catalán. (Castaño Muñoz, 2011)

Descripción: Esta investigación trata la importancia de estudiar algunos elementos sobre el uso del internet para el proceso de enseñanza y aprendizaje. Se centra principalmente en el análisis identificado como: el uso de internet para el aprendizaje interactivo. Su objetivo principal es comparar la eficacia de la utilización de internet en educación, para el aprendizaje interactivo con la del uso de internet en educación para el aprendizaje individual. (Castaño Muñoz, 2011), de donde se derivan sus objetivos específicos.

- Estudiar la forma de la productividad del uso de internet para la interacción en el proceso de enseñanza-aprendizaje.
- Estudiar los efectos heterogéneos del uso de internet para el aprendizaje interactivo en el rendimiento académico entre diferentes grupos de alumnos usuarios de internet.
- Explorar cuáles son los determinantes que facilitan a los alumnos usuarios de internet hacer uso de internet para el aprendizaje interactivo.

Las variables que se tomaron en cuenta en el proyecto son: variables observadas directamente: género y edad, brecha digital, variables académicas, usos de Internet para la educación, variables no observadas directamente, tiempo de estudio disponible.

Técnicas empleadas: Se utilizan las técnicas multivariadas, para el proceso más explicativo se basa principalmente en regresiones lineales múltiples.

2.6.2. Proyecto: Relación entre el uso de internet y el logro académico (Ayala, 2007)

Descripción: Esta investigación estudia las relaciones que existen entre los usos de internet con el logro académico de los estudiantes de Chile. Se analiza la existencia de brecha digital en varios grupos socioeconómicos y de cómo utilizan el internet en actividades que se reporta una relación significativa con el logro educacional. (Ayala, 2007).

Las variables utilizadas son: variables dependientes, educación de padre, sexo, experiencia, frecuencia de uso de internet, rendimiento académico, variables independientes, conjunto de preguntas de la Encuesta WIP 2006.

Técnicas empleadas: Las técnicas utilizadas en este proyecto son el análisis multivariado, tales como análisis factorial, regresión logística. El software utilizado es Stata.

2.6.3. Proyecto: 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 Díaz, 2012)

Descripción: El presente proyecto se centra principalmente en explorar las relaciones y niveles de incidencia de las variables socio-demográficas, nivel de ingreso, edad y género del estudiante sobre las clasificaciones mencionadas, y, en su parte medular a explicar el efecto del uso de internet de estudiantes y profesores sobre el rendimiento académico.(Torres Díaz, 2012). Entre sus objetivos están

- Explorar las tipologías de usuarios subyacentes en función de los usos de internet, el acceso a dispositivos tecnológicos y percepciones respecto a la confianza e importancia de internet para la actividad académica.
- Determinar el nivel de incidencia de las variables socio-demográficas sobre las tipologías de usuarios.
- Determinar el nivel de incidencia de los usos de internet sobre el rendimiento académico.
- Establecer la relación existente entre el uso de internet en el aula por parte del profesor y el rendimiento académico.

Las variables utilizadas son: edad, género, nivel de ingresos, lugar de conexión, rendimiento académico.

Técnicas empleadas: Se utilizan las técnicas estadísticas: análisis factorial, análisis cluster, análisis de varianza, análisis discriminante y regresión logística. El software utilizado para realizar los procesos IBM SPSS Statistics Base 19.

2.6.4. Proyecto: Las técnicas de Clustering en la personalización de sistemas de e-learning (Farias, Durán, & Figueroa, 2008).

Descripción: Este trabajo se centra a descubrir información sobre el estilo de aprendizaje dominante del estudiante, con base al análisis de Cluster de las interacciones que realiza en un entorno de aprendizaje virtual. Este trabajo se enfoca principalmente de dotar a un sistema de e-learning de las herramientas y estrategias necesarias para evitar la deserción de estudiante en los cursos regulares.

- permitir que los profesores definan de una forma muy sencilla un test adaptativo informatizado.
- permitir que los alumnos realicen los test definidos y sean evaluados por el sistema, todo ello a través de la Web.

Técnicas empleadas: Técnicas de Análisis de Cluster.

Así mismo existen más trabajos referentes a la temática del presente investigación, lo podemos encontrar en (Álvarez Sáiz & Zorrilla Pantaleón, 2012), en este estudio se analiza profundamente la relación entre los estilos de aprendizaje, el rendimiento académico y la actividad en las plataformas virtuales de estudiantes de la universidad de Cantabria. Este estudio indica como resultados, que el estilo de aprendizaje utilizando las nuevas tecnologías 2.0 así como el uso de internet, no es un factor influyente en el rendimiento.

En (Rodríguez Morales, 2015), además de obtener información sobre los perfiles y comportamientos en el aprendizaje de sus alumnos, este trabajo ha podido determinar cuáles son las variables más influyentes en el desempeño académico del alumno y a partir de esta, posibilita generar reglas que permiten tomar decisiones en el momento adecuado para influir positivamente al estudiante en sus estudios.

CAPITULO III: METODOLOGÍA

Al ser este un proyecto de investigación nacional, a través del director del proyecto se proporcionó los aspectos metodológicos a utilizar.

- Población y muestra.
- Aplicación de la encuesta “Usos de tecnología en las universidades” basado en los utilizados en los proyectos PIC, DLINHE, ECUADOR.
- Aplicación de análisis multivariante.
- Aplicación de regresiones.

3.1. Población y muestra

La población objetivo, son los estudiantes de la Universidad Técnica Particular de Loja, a partir del tercer ciclo.

En lo que corresponde a la determinación del tamaño muestral, se ha utilizado el muestreo probabilístico, ya que al ser parte de una población finita, el nivel de confianza es de 95% con un margen de error del 5%. Para maximizar el tamaño de la muestra utilizamos el valor de $p=q=50$. Para el cálculo del tamaño de la muestra de esta investigación partimos de la formulación para poblaciones finitas propuesta por (Arnal, Rincón, & Latorre, 1992).

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

En donde:

n = tamaño de la muestra.

N =valor de población.

Z = Valor crítico correspondiente un coeficiente de confianza del cual se desea hacer la investigación.

p = Proporción proporcional de ocurrencia de un evento.

q = Proporción proporcional de no ocurrencia de un evento

d = equivale al error aceptado.

Remplazando los valores:

$N=6.073$

$Z= 0,95= 1,96 \%$

$p = 50 =0,5$

$q = 50=0,5$

$d = 0,05 (5\%)$

Sustituimos la formula anterior con los valores dados, quedándonos como tamaño de la muestra la siguiente:

$$n = \frac{1,96^2 * 0,5 * 0,5 * 6073}{0,05^2 * (6073 - 1) + 1,96^2 * 0,5 * 0,5}$$

$$n=361,360$$

$$n= \mathbf{361}$$

La muestra estuvo formada por 500 estudiantes, luego de ser tabuladas y filtradas se procedió a revisar los datos atípicos, a fin de evitar distorsiones en los resultados, no obstante se quedó con 451 encuestas, con lo cual se puede certificar que la muestra con la que se trabajó en esta investigación es superior a la solicitada.

3.2. Recolección, Procesamiento y Análisis de datos

3.2.1. Recolección de Información

Para desarrollar esta investigación, se requirió la aplicación de un conjunto de estrategias o procedimientos, para obtener los resultados de los objetivos trazados, y la demostración de las hipótesis. Las técnicas que se utilizó para realizar esta investigación fueron la técnica de la encuesta y la entrevista.

Se aplicó la encuesta proporcionada por el director del proyecto, “Usos de tecnología en las universidades” basado en los utilizados en los proyectos PIC, DLINHE, ECUADOR en donde se emplearon 17 ítems dirigidos a explorar sobre las variables sociodemográficas y los diversos aspectos relacionados con el uso de internet en el ámbito académico y de entretenimiento. Ver Anexo 1.

A continuación se detallan las variables que se utilizaron para el análisis de la investigación:

- Primer bloque: Universidad, titulación del estudiante y variables sociodemográficos. Ver Tabla 1.

Tabla 1: Descripción de las variables sociodemográficas

Campo	Variable	Tipo de Dato	Descripción	Valores
Uni	Nominal	Cadena	Universidad del estudiante	Ninguna
Car	Nominal	Cadena	Titulación del estudiante	Ninguna
Eda	Real	Numérico	Edad del estudiante	Ninguna
Sex	Dicotómica	Numérico	Genero del estudiante	1= Hombre; 2= Mujer
Ing	Nominal	Numérico	Ingresos económicos mensuales de la familia del estudiante	1=Hasta 350;2=Hasta 600; 3=Hasta 1.000; 4=Hasta 1.500; 5=Más de 1.500 dólares

Fuente: Elaboración propia

- Segundo bloque: Preguntas sobre el conocimiento y preferencias del estudiante para conectarse desde el internet; igualmente se indagó sobre las horas y años que navega en Internet. Ver Tabla 2

Tabla 2: Descripción de las variables de conocimiento y preferencias del estudiante en el uso de Internet.

Campo	Variable	Tipo de Dato	Descripción	Valores
lug_con	Nominal	Numérico	Lugar desde dónde se conecta habitualmente a Internet el estudiante	1=Desde la casa; 2=Desde un cyber café; 3=Desde el trabajo; 4=Desde la Universidad; 5=Desde una red móvil
dia_con	Ordinal	Numérico	días a la semana se conecta Internet el estudiante	1=7
niv_con	Ordinal	Numérico	Nivel de conocimiento en el manejo de Internet del estudiante	1=10
hor_dia	Nominal	Numérico	Horas que es estudiante se conecta al Internet por día	Ninguna
año_exp	Nominal	Numérico	Hace cuántos años se conecta a Internet.	Ninguna

Fuente: Elaboración propia

- Tercer bloque: Cuerpo central del cuestionario, aquí se profundiza en el uso de internet, se exploró información sobre actividades académicas mediante 10 variables descrita en la Tabla 3.

Tabla 3: Descripción de las variables referentes al uso de Internet Académico.

Campo	Variable	Tipo de Dato	Descripción	Valores
a_ing_pla	Nominal	Numérico	Cuántas veces a la semana ingresa a la plataforma virtual de su universidad el estudiante.	Ninguna
a_con_pro	Nominal	Numérico	Consultas que haces los estudiantes a sus profesores cada mes sobre temas académicos.	Ninguna
a_con_est	Nominal	Numérico	Consultas que el estudiante realiza a sus compañeros por mes sobre temas académicos.	Ninguna
a_rec_edu	Nominal	Numérico	Recursos educativos que el estudiante descarga de la plataforma virtual cada mes.	Ninguna
a_vid_aca	Nominal	Numérico	Videos académicos que el estudiante mira en YouTube cada mes.	Ninguna
a_for_vir	Nominal	Numérico	Foros virtuales que el estudiante participa cada mes sobre temas académicos.	Ninguna
a_post_aca	Nominal	Numérico	Post o tweets que el estudiante realiza sobre temas académicos en las redes sociales por mes.	Ninguna
a_hor_cha	Nominal	Numérico	Horas que el estudiante chatea sobre temas académicos cada mes.	Ninguna
a_bus_inf	Nominal	Numérico	Horas que el estudiante busca información académica en internet cada mes.	Ninguna
a_bib_vir	Nominal	Numérico	Horas que el estudiante utiliza la biblioteca virtual de la universidad.	Ninguna

Fuente: Elaboración propia

- Cuarto bloque: Se preguntó acerca de lo referente al uso que los estudiantes le dan al internet en el ámbito de entretenimiento y diversión. Ver Tabla 4

Tabla 4: Descripción de las variables referentes al uso de Internet en entretenimiento y diversión.

Campo	Variable	Tipo de Dato	Descripción	Valores
e_hor_cha	nominal	Numérico	Horas a la semana que el estudiante chatea por diversión.	Ninguna
e_red_soc	nominal	Numérico	Horas a la semana que el estudiante utiliza redes sociales.	Ninguna
e_jue_lin	nominal	Numérico	Horas a la semana que el estudiante utiliza juegos en línea.	Ninguna
e_des_con	nominal	Numérico	Horas a la semana que el estudiante descarga música, videos y programas.	Ninguna
e_vid_sem	nominal	Numérico	Videos que el estudiante mira en YouTube cada semana para entretenimiento.	Ninguna
rs_seg_twi	nominal	Numérico	Seguidores que tiene en Twitter el estudiante.	Ninguna
rs_seg_fac	nominal	Numérico	Amigos de Facebook del estudiante.	Ninguna
rs_seg_lin	nominal	Numérico	Contactos que tiene en LinkedIn el estudiante.	Ninguna
cc_blog	dicotómica	Numérico	Si el estudiante tiene un blog.	1= Si; 2= No
cc_youtube	dicotómica	Numérico	Si el estudiante tiene cuenta en YouTube.	1= Si; 2= No
cc_delicious	dicotómica	Numérico	Si el estudiante tiene cuenta en www.del.icio.us.	1= Si; 2= No

Fuente: Elaboración propia

- Y finalmente, se realizó preguntas de las percepciones que los estudiantes tienen acerca del nivel de uso de dispositivos electrónicos; Así mismo, se indagó sobre las asignaturas que se matriculo y que a su vez reprobó. Ver Tabla 5.

Tabla 5: Descripción de las variables; nivel de uso de dispositivos y materias que el estudiante se matriculó.

Campo	Variable	Tipo de Dato	Descripción	Valores
d_tel_cam_int	Ordinal	Numérico	Nivel de uso del estudiante, Smartphone con cámara fotográfica y acceso a internet	1=10
d_tel_int	Ordinal	Numérico	Nivel de uso del estudiante, Teléfono móvil con acceso a internet	1=10
d_tel	Ordinal	Numérico	Nivel de uso del estudiante, Teléfono móvil sin acceso a internet	1=10
d_com_por	Ordinal	Numérico	Nivel de uso del estudiante, Computador portátil	1=10
d_tab	Ordinal	Numérico	Nivel de uso del estudiante, Tablet (iPad; Galaxy Tab, Kindle, etc)	1=10
d_cam_dig	Ordinal	Numérico	Nivel de uso del estudiante, Cámara digital	1=10
d_ipod	Ordinal	Numérico	Nivel de uso del estudiante, iPod / MP3 Player	1=10
cic_asi_mat	Nominal	Numérico	Asignaturas que se matriculó el estudiante en el ciclo anterior.	Ninguna
cic_asi_apr	Nominal	Numérico	Asignaturas que aprobó el estudiante en el ciclo anterior.	Ninguna

Fuente: Elaboración propia

3.2.2. Procesamiento y Análisis de Datos

Para realizar el procedimiento se utilizó los métodos estadísticos de análisis multivariante, en los cuales se analizan las relaciones entre diversas variables independientes y una variable dependiente. Las técnicas estadísticas multivariadas empleadas en esta investigación son: análisis de conglomerados (cluster), análisis discriminante y regresión logística multinomial.

El siguiente método es el aplicado para la clasificación de los estudiantes en grupos de acuerdo al uso que le dan al internet.

1. Agrupar a los estudiantes mediante análisis de conglomerados (cluster).
2. Verificar el cluster mediante análisis discriminante.
3. Elaboración de tablas de doble entrada, para aplicar el estadístico de chi cuadrado. (X^2).
4. Comprobación de hipótesis mediante regresión logística multinomial.

3.2.3. Agrupar a los estudiantes mediante análisis de Conglomerados

Para esta investigación se utilizó el análisis cluster para agrupar las variables que conciernen a las actividades y preferencias académicas y de entretenimiento. Para establecer a los estudiantes por grupos se utilizó el programa SPSS versión 17. Se realizó la

clasificación con tres, cuatro y cinco cluster (ver anexo 2), para garantizar una correcta agrupación.

3.2.4. Verificar el cluster mediante análisis Discriminante

Como se mencionó anteriormente se realizó la agrupación a los estudiantes en 3, 4 y 5 clusteres. Se procedió a realizar el análisis discriminante para verificar el porcentaje de exactitud más exacto. Ver Anexo 3 y Anexo 6.

3.2.5. Elaboración de tablas de doble entrada, para aplicar el estadístico de Chi cuadrado. (χ^2)

Después con el grupo que se escoge para continuar con la investigación se procedió a realizar una relación (tabla de doble entrada) entre las variables ingresos (ing) y los usos de internet en actividades académicas (clus_aca2) y de la variables ingresos (ing) y los usos de internet en actividades de entretenimiento (clus_entre2), se utilizó la técnica estadística chi-cuadrado de Pearson, ya que permite verificar la existencia de una relación entre dos variables tipo nominal.

Para la presente investigación se trabajó con un nivel de confianza de 95%. Entonces se procedió a la determinación de la hipótesis nula y alternativa:

- H0: No hay relación entre las variables
- H1: Si hay relación entre las variables

El valor de p indica que se alcanzado el umbral de significación estadística, en este caso se toma por convenio el límite de 0,05, de forma que si p vale más de ese valor se considera que no existe relación alguna entre las variables. Es decir, se rechaza la H1 y se acepta la H0.

3.2.6. Comprobación de Hipótesis mediante Regresión Logística.

Como las variables dependientes tienen más de dos categorías, se utilizó la técnica de regresión logística multinomial, ya que considera variables dependientes de tipo nominal con más de dos categorías, y que permite predecir las probabilidades de los diferentes resultados posibles.

Para realizar la comprobación de la primera hipótesis planteada en esta investigación, que sostiene que el nivel de ingresos económicos de los estudiantes determina como se utiliza

internet para el aprendizaje, se considera las siguientes variables, (clus_aca2), que cuenta con tres categorías que son Destacado=1; Intermedio=2; Tradicional=3 y la variable ingresos, considerando los techos de: 350 =1, 600=2, 1.000=3; 1.500 =4 y Más de 1.500 dólares.

Para realizar la comprobación de la segunda hipótesis planteada en esta investigación que considera que el nivel de ingresos económicos de los estudiantes determina como se utiliza internet para entretenimiento se considera las variables (clus_entre2), que cuenta con tres categorías Motivado=1; Pasivo=2 y Tradicional=3 y la variable ingresos, tomando en cuenta los siguientes márgenes: 350 =1, 600=2, 1.000=3; 1.500 =4 y Más de 1.500 dólares.

Se utiliza la prueba de Wald para verificar los coeficientes de las variables independientes y el test del chi cuadrado.

Para la comprobación de las hipótesis 3 y 4 se utiliza la regresión logística binaria, ya que la variable dependiente es una variable binaria o también conocida como dummy o dicotómica.

3.3. Metodología para el uso de la herramienta Weka.

Para efectos del presente trabajo, la base de datos que se va tratar corresponde a la misma que se utilizó para el análisis mediante la Herramienta SPSS. El proceso que se va a seguir es el KDD (descubrimiento de conocimiento en base de datos de (Hernández et al., 2007), tal como lo indica la Figura 7.

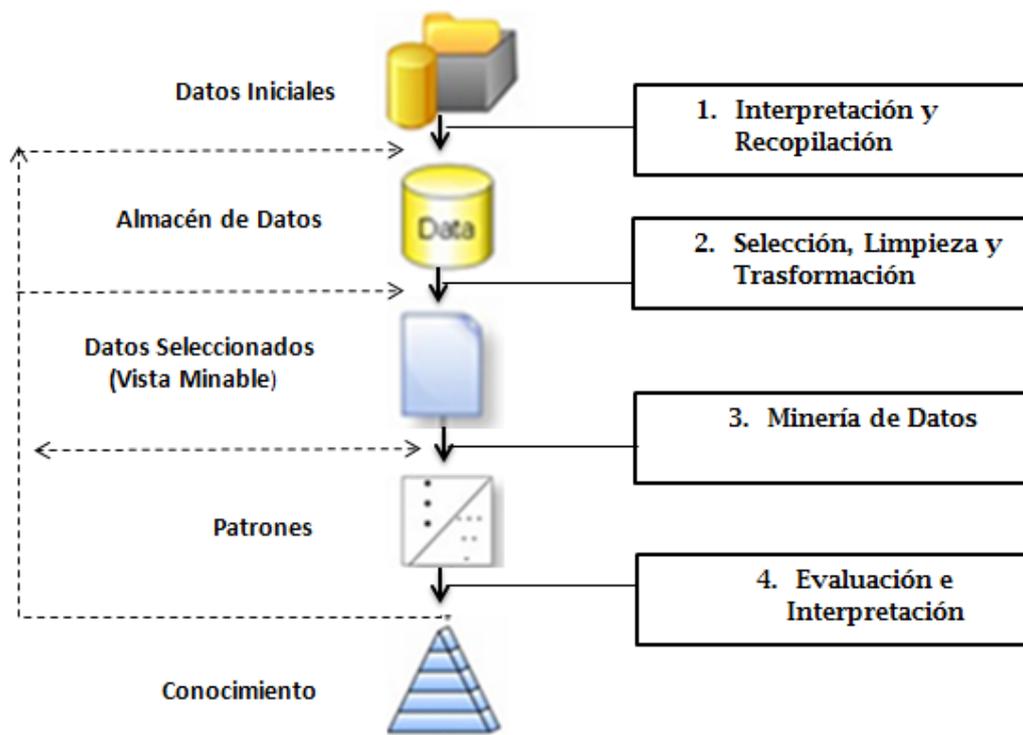


Figura 7: Fases del proceso de extracción del conocimiento

Fuente:(Hernández et al., 2007)

3.3.1 Integración y recopilación

Los datos necesarios para la presente fase provienen de una encuesta aplicada a estudiantes de la universidad, esta encuesta ha sido tabulada y procesada para propósitos de la investigación.

En la sección 3.2.1 se puede observar el procedimiento para la recopilación de la información.

3.3.2 Selección, limpieza y transformación de los datos

A continuación en esta fase, se construirá ya el DataSet con los registros que se ingresaran a la herramienta Weka, ya sin datos atípicos.

El objetivo de esta investigación es determinar el nivel de incidencia de los Usos de Internet con los ingresos económicos de las familias de los estudiantes, y a su vez medir la incidencia de éstos con el rendimiento académico; es por eso que de la Base de datos nos interesó todas aquellas variables que tengan relación con las hipótesis planteadas en el inicio del presente trabajo y desde las cuales se procederá a determinar agrupaciones y patrones sobre estos grupos.

- Para las Hipótesis uno y tres: Variables Involucradas, aprendizaje, rendimiento académico e ingresos del núcleo familiar. El detalle completo se encuentra en la siguiente Tabla 6.

Tabla 6: Variables consideradas para el análisis de Hipótesis 1 y 3

Variabl e	Tipo	Long itud	Descripción	Valor por Defecto
a_ing_pl a	Numérico	8	¿Cuántas veces a la semana ingresa a la plataforma virtual de su universidad?	numeric
a_con_p ro	Numérico	8	Aproximadamente cuántas consultas les hace a sus profesores cada mes	numeric
a_con_e st	Numérico	8	Aproximadamente cuántas consultas realiza a sus compañeros por mes	numeric
a_rec_e du	Numérico	8	Aproximadamente cuántos recursos educativos descarga de la plataforma virtual cada mes	numeric
a_vid_ac a	Numérico	8	Aproximadamente cuántos videos académicos mira en youtube cada mes	numeric
a_for_vir	Numérico	8	Aproximadamente en cuántos foros virtuales participa cada mes	numeric
a_post_a ca	Numérico	8	Aproximadamente cuántos post o tweets sobre temas académicos realiza en las redes sociales por mes	numeric
a_hor_c ha	Numérico	8	Aproximadamente cuántas horas chatea sobre temas académicos cada mes	numeric
a_bus_in f	Numérico	8	Aproximadamente cuántas horas busca información académica en internet cada mes	numeric
a_bib_vir	Numérico	8	Aproximadamente cuántas horas utiliza la biblioteca virtual de la universidad cada mes	Numeric
Ing	Numérico	8	Ingresos económicos mensuales de la familia del estudiante	1=Hasta 350;2=Hasta 600; 3=Hasta 1.000; 4=Hasta 1.500; 5=Más de 1.500 dólares
rendimie nto_acad	numérico	8	Rendimiento académico del estudiante	1=Aprobado;2=Reprobado

Fuente: Elaboración propia

- Para las hipótesis 2 y 4 se consideró todas las variables relacionadas con el entretenimiento, el detalle de todas ellas se encuentran en la Tabla 7.

Tabla 7: Variables consideradas para el análisis de Hipótesis 2 y 4

Variable	Tipo	Longitud	Descripción	Valor por Defecto
e_hor_cha	Numérico	8	Horas a la semana que el estudiante chatea por diversión.	numeric
e_red_soc	Numérico	8	Aproximadamente cuántas horas a la semana utiliza redes sociales	numeric
e_jue_lin	Numérico	8	Aproximadamente cuántas horas a la semana utiliza juegos en línea	numeric
e_des_con	Numérico	8	Aproximadamente cuántas horas a la semana descarga música, videos y programas	numeric
e_vid_sem	Numérico	8	Aproximadamente cuántos videos para entretenimiento mira en YouTube cada semana	numeric
Ing	Numérico	8	Ingresos económicos mensuales de la familia del estudiante	1=Hasta 350;2=Hasta 600; 3=Hasta 1.000; 4=Hasta 1.500; 5=Más de 1.500 dólares
rendimient_o_acad	numérico	8	Rendimiento académico del estudiante	1=Aprobado;2=Reprobado

Fuente: Elaboración propia

Continuando, una vez ya seleccionadas las variables se procedió a considerar lo siguiente: como los valores de las variables seleccionadas, no proporcionaban información específica relevante, fue necesario normalizar esta variable de tipo académico en rangos de horas, se realizó una nueva variable, o también conocido como agregación¹, los cuales se detalla a continuación:

- Para la construcción de la nueva variable, se procedió a recodificar la variable niv_con. Ver Tabla 8 con ayuda de una hoja de cálculo Excel. Ver Anexo 14.

Tabla 8: Variable a ser recodificada para medir los usos de Internet

Campo	Variable	Tipo de Dato	Descripción	Valores
niv_con	Ordinal	Numérico	Nivel de conocimiento en el manejo de Internet del estudiante	1=10

Fuente: Elaboración propia

- Se fijaron unos umbrales² o rangos mínimos de soporte y confianza, quedando de la siguiente manera.

¹Agregación.- Consiste en crear nuevos atributos para mejorar la calidad, visualización o comprensibilidad del conocimiento extraído. La mayoría o todos los atributos originales se preservan, añadir nuevos atributos y no de sustituirlos.(Hernández et al., 2007)

² Umbral.- Permiten encontrar reglas $X \& Y \Rightarrow Z$ con mínimo % de confianza y soporte, mejoran la eficiencia del algoritmo. también llamados puntos de corte.

Tabla 9: Umbrales para la construcción de la variable usos_internet_academico

Variable Construida	Umbral o rango	Descripción
uso_internet_academico	6	1-6 = menor uso de Internet 7-10= mayor uso de Internet
	5	1-5= menor uso de Internet 6-10= mayor uso de Internet
	4	1-4 = menor uso de Internet 5-10= mayor uso de Internet
	3	1-3 = menor uso de Internet 4-10= mayor uso de Internet
	2	1-2 = menor uso de Internet 3-10= mayor uso de Internet

Fuente: Elaboración propia

Seguidamente, se realizó la exportación de la base de datos de Excel, al archivo arff que es el formato oficial de la herramienta Weka.

3.3.3 Minería de Datos y extracción de patrones

En esta fase, se procedió aplicar de las técnicas de minería de datos seleccionadas. Con base en los objetivos planteados para esta investigación y considerando varios estudios realizados en esta línea, la técnica que más se adaptó es el algoritmo de K-means, ya que nos permite agrupar a los estudiantes en subclases de acuerdo a su nivel de uso de Internet tanto en el ámbito académico como de entretenimiento.

El modelo generado que describió los patrones y relaciones entre los datos es:

1. Algoritmos para selección de atributos,
 - Evaluador de conjunto de atributos: **CfsSubsetEval**
 - Método de búsqueda: **GreedyStepwise**
2. Algoritmo de agrupamiento
 - K-means

3.3.3.1 Algoritmos para selección de atributos

Para mejorar la calidad del modelo se procedió aplicar algoritmos para selección de atributos, así mismo, estos algoritmos se encargan de eliminar los datos irrelevantes o redundantes.

3.3.3.1.1 DataSet para la hipótesis uno

Los algoritmos con los que mejores resultados que se obtuvieron fueron con: CfsSubsetEval como evaluador de atributos y GreedyStepwise como método de búsqueda. Experimentos ver Anexo 15

Después de realizar varios experimentos se pudo obtener que las variables de: **a_post_aca** y **a_bus_inf** son las variables más representativas ya que se encuentran presente en todos los grupos formados por los algoritmos evaluadores.

Con respecto a la variable **a_for_vir**, aunque no estuvo presente en la mayoría de grupos formados por los algoritmos evaluadores tuvo en rendimiento creciente por lo cual también fue considerada dentro del análisis. El detalle completo de los experimentos se los puede observar en la Tabla 10. Ver Anexo 16.

Tabla 10: Resultados obtenidos de los algoritmos de selección de atributos para variables académicas

Attribute	Experimento 1	Experimento 2	Experimento 3	Experimento 4	Experimento 5
	%	%	%	%	%
a_con_pro	60	90	87	100	100
a_con_est	60	80	67	65	84
a_for_vir	80	90	93	100	100
a_post_aca	100	100	100	100	100
a_bus_inf	100	100	100	100	100
a_ing_pla	N/A	60	67	85	80
a_vid_aca	N/A	70	67	65	80

Fuente: Elaboración propia.

El DataSet³ que se obtuvo después de los experimentos con el cual se trabajara para la comprobación de la Hipótesis se lo muestra en la Tabla 11.

³ **DataSet.**- Un "Conjunto de datos" o "dataset" es una colección de datos normalmente tabulada.

Tabla 11: DataSet para la comprobación de la Hipótesis uno, con Weka

Variable	Tipo	Longitud	Descripción	Valor por Defecto
a_for_vir	Numérico	8	Aproximadamente en cuántos foros virtuales participa cada mes	Numeric
a_post_ac a	Numérico	8	Aproximadamente cuántos post o tweets sobre temas académicos realiza en las redes sociales por mes	Numeric
a_bus_inf	Numérico	8	Aproximadamente cuántas horas busca información académica en internet cada mes	Numeric
Ing	Numérico	8	Ingresos económicos mensuales de la familia del estudiante	1=Hasta 350;2=Hasta 600; 3=Hasta 1.000; 4=Hasta 1.500; 5=Más de 1.500 dólares
Uso_intern et_acade mico_4	Numérico		Clase, uso de Internet del estudiante	1 = menor uso de Internet 2= mayor uso de Internet

Fuente: Elaboración propia.

3.3.3.1.2 DataSet para la hipótesis dos

De forma similar se procedió aplicar los algoritmos de selección de atributos, (CfsSubsetEval y GreedyStepwise) para determinar las variables más representativas para el análisis, y los resultados que se obtuvieron fueron los siguientes: Tabla 12.

Tabla 12: Resultados de selección de atributos para variables de entretenimiento

Attribute	Experimento 1	Experimento 2	Experimento 3	Experimento 4	Experimento 5
	%	%	%	%	%
e_hor_cha	100	100	100	100	100
e_red_soc	0	0	0	0	0
e_jue_lin	20	10	7	5	4
e_des_con	80	100	100	100	100
e_vid_sem	80	80	60	70	56

Fuente: Elaboración propia

Como se puede observar solo la variable: **e_hor_cha**, está presente en todas las agrupaciones hechas por los algoritmos, mientras que **e_des_con**, logra obtener el 100 % de representatividad en 4 de 5 experimentos realizados.

Adicionalmente se consideró a la variable **e_vid_sem** también en el análisis ya que sus resultados son los más altos comparados con las variables restantes.

A continuación en la Tabla 13 se detallan los atributos del DataSet, que se utilizará para la generación del modelo.

Tabla 13: DataSet para la comprobación de la Hipótesis dos, con Weka

Variable	Tipo	Longitud	Descripción	Valor por Defecto
e_hor_cha	Numérico	8	Horas a la semana que el estudiante chatea por diversión.	numeric
e_des_con	Numérico	8	Aproximadamente cuántas horas a la semana descarga música, videos y programas	numeric
e_vid_sem	Numérico	8	Aproximadamente cuántos videos para entretenimiento mira en YouTube cada semana	Numeric
Ing	Numérico	8	Ingresos económicos mensuales de la familia del estudiante	1=Hasta 350;2=Hasta 600; 3=Hasta 1.000; 4=Hasta 1.500; 5=Más de 1.500 dólares

Fuente: Elaboración propia

3.3.3.1.3 DataSet para la hipótesis tres

La hipótesis tres sostiene, que el uso de internet en el aprendizaje incide en el rendimiento académico. Es por esta razón que las variables que nos interesan son todas aquellas que intervienen en ésta, como en el apartado 3.1.2.1.1 ya se realizó el proceso de selección de variables del uso de Internet académico, utilizaremos estas mismas para el DataSet y para el rendimiento académico se tomó la variable del apartado 4.5. Quedando el siguiente DataSet. Ver Tabla 14.

Tabla 14: DataSet para la comprobación de la Hipótesis tres, con Weka

Variable	Tipo	Longitud	Descripción	Valor por Defecto
a_for_vir	Numérico	8	Aproximadamente en cuántos foros virtuales participa cada mes	numeric
a_post_aca	Numérico	8	Aproximadamente cuántos post o tweets sobre temas académicos realiza en las redes sociales por mes	numeric
a_bus_inf	Numérico	8	Aproximadamente cuántas horas busca información académica en internet cada mes	numeric
rendimiento_acad	numérico	8	Rendimiento académico del estudiante	1=Aprobado; 2=Reprobado

Fuente: Elaboración propia

3.3.3.1.4 DataSet para la hipótesis cuatro

De forma similar, se procedió a optar por las variables resultantes del apartado 3.1.2.1.2 aumentado la variable de rendimiento académico. Tal como lo muestra la siguiente Tabla 15.

Tabla 15: DataSet para la comprobación de la Hipótesis cuatro, con Weka

Variable	Tipo	Longitud	Descripción	Valor por Defecto
e_hor_cha	Numérico	8	Horas a la semana que el estudiante chatea por diversión.	numeric
e_des_con	Numérico	8	Aproximadamente cuántas horas a la semana descarga música, videos y programas	numeric
e_vid_sem	Numérico	8	Aproximadamente cuántos videos para entretenimiento mira en YouTube cada semana	Numeric
rendimiento_acad	Numérico	8	Rendimiento académico del estudiante	1=Aprobado 2=Reprobado

Fuente: Elaboración propia

3.3.3.2 Algoritmo de agrupación (K-means)

Como ya se mencionó anteriormente, la técnica de agrupamiento o clustering que mejor se adaptó a los propósitos del presente trabajo es el algoritmo de *SimpleKMeans*. Para aplicar este algoritmo es necesario realizar algunas configuraciones de los valores que están establecidos por defecto en Weka. Ver Tabla 16.

Tabla 16: DataSet Configuración de opciones en el algoritmo SimpleKMeans.

Parámetro	Descripción	Valor
displayStdDevs	Esta función sirve para mostrar el valor de la división estándar de los valores numéricos	False
distanceFunction	Calcula las distancias entre los valores, dejamos la que viene por defecto.	Euclidean distance (or similarity) function.
maxIterations	Número máximo de iteraciones del algoritmo si esto no convergió antes.	500
numClusters	Establece en número de clusteres que se desea.	2
Seed	o semilla, para elegir el este número se realiza varios experimentos, tomando en cuenta el valor que muestre el menor error cuadrático.	Hipótesis 1=20 Seed Hipótesis 2= Hipótesis 3=25 Seed Hipótesis 4=

Fuente: Elaboración propia

- **Resultados de los experimentos de agrupamiento**

Continuamente ya teniendo los DataSet se procedió a realizar los experimentos con los diferentes umbrales y con varios Seed con el fin de obtener el menor error cuadrático. Ver Anexo 16.

a) Clusters de la Hipótesis 1

Para obtener los resultados de esta hipótesis se realizaron cinco experimentos, se concluyó que la configuración ideal del algoritmo *SimpleKMeans*, corresponde a un Seed de 20 a 25. Con el cual se ha obtenido el mínimo número de error cuadrático de los experimentos realizados. Ver Tabla 17.

Tabla 17: DataSet Resultados de experimentos para la hipótesis 1

Experimento	Seed	Sum of squared errors	Raíz
1	10	260,490064	16,1397046
2	15	269,88222	16,4280924
3	20	121,677974	11,0307739
4	25	121,677974	11,0307739
5	30	221,677974	14,888854

Fuente: Elaboración propia

b) Clusters de la Hipótesis 2

Para esta hipótesis se realizó dos análisis:

- En el primero se utilizaron aquellas variables que fueron seleccionadas utilizando algoritmos de selección de atributos. Ver Tabla 18

Tabla 18: DataSet Resultados de experimentos para la hipótesis 2

Experimento	Seed	Sum of squared errors	Raíz
1	10	87,1904	9,337580618
2	15	87,1904118	9,337580618
3	20	88,9283	9,430182378
4	25	87,1898	9,337546985
5	30	88,9283	9,430182378

Fuente: Elaboración propia

Sin embargo, las variables obtenidas en la sección anterior de selección de atributos resultaron no ser relevantes para encontrar patrones útiles con el algoritmo de agrupamiento.

Se procedió a realizar un segundo análisis, esta vez con todas las variables que interviene en el uso de Internet para entretenimiento. Ver Tabla 19.

Tabla 19: DataSet Resultados del segundo análisis de experimentos para la hipótesis 2

Experimento	Seed	Sum of squared errors	Raíz
1	10	119,7894	10,9448365
2	15	119,7894	10,9448365
3	20	123,3592	11,1067206
4	25	119,7894	10,944833
5	30	123,3592	11,1067206

Fuente: Elaboración propia

c) Clusters de la Hipótesis 3

De todos los experimentos realizados. (Ver tabla 50) el número 4 tiene la menor suma de errores cuadráticos: 51,5824, este experimento utiliza un Seed de 25. Con el que se procederá a trabajar y describir más adelante en el capítulo de resultados. Ver Anexo 17.

Tabla 20: DataSet Resultados de experimentos para la hipótesis 3

Experimento	Seed	Sum of squared errors	Raíz	Instancias incorrectas
1	10	51,5826	7,18210324	185,0000
2	15	51,6585	7,18738313	165,0000
3	20	51,5826	7,18210324	185,0000
4	25	51,5824	7,18209044	184,0000
5	30	51,6157	7,18440941	167,0000

Fuente: Elaboración propia

d) Clusters de la Hipótesis 4

El análisis de la hipótesis se la realizó de dos formas:

- En la primera se consideraron aquellas variables obtenidas de la aplicación de filtros de la sección 3.3.2, obteniéndose las siguientes variables como relevantes para el estudio, Sin embargo, en los experimentos realizados no se presentan evidentes características entre los dos grupos formados por el algoritmo ya que información relacionada a las variables de análisis no tienen mayores diferencias entre los dos grupos. Ver Tabla 21.

Tabla 21: DataSet Resultados de experimentos para la hipótesis 4

Experimento	Seed	Sum of squared errors	Raíz	Incorrectly clustered instances
1	10	77,2689	8,790270497	41,4634
2	15	77,2689	8,790270497	41,4634
3	20	77,2669	8,79015706	41,9069
4	25	77,2669	8,79015706	41,9069
5	30	77,2683	8,790238836	41,6851

Fuente: Elaboración propia

En virtud de esto se procedió a realizar un segundo análisis Ver Anexo 17, en donde se involucraban todas las variables que tienen que ver con actividades de entretenimiento de los estudiantes. Ver Tabla 22.

Con estos resultados se procederá a trabajar con el experimento uno. Y se lo explicara en la sección de resultados.

Tabla 22: DataSet Resultados del segundo análisis de experimentos para la hipótesis 4

Experimento	Seed	Sum of squared errors	Raíz
1	10	75,5090	8,68959052
2	15	150,4422	12,26549
3	20	75,5090	8,68959052
4	25	150,4464	12,265658
5	30	150,4464	12,265658

Fuente: Elaboración propia

3.3.4 Evaluación e Interpretación

En esta fase se analizaran los resultados de cada una de las experimentaciones seleccionadas y se evaluará las posibles soluciones.

CAPÍTULO IV: RESULTADOS

4.1 Descripción de la muestra

Se presentan a continuación los resultados obtenidos de las encuestas realizadas a una muestra de 451 estudiantes lo cual permitirán conocer los factores que influyen en el uso del internet en el aprovechamiento académico en la Universidad Técnica Particular de Loja, asimismo permitirá relacionar los usos de internet con los niveles de ingreso de los estudiantes.

De los estudiantes encuestados la mayor parte de la muestra estuvo formada por mujeres (55%) el restante 45% corresponde a los hombres, tal como lo muestra la Figura 8.

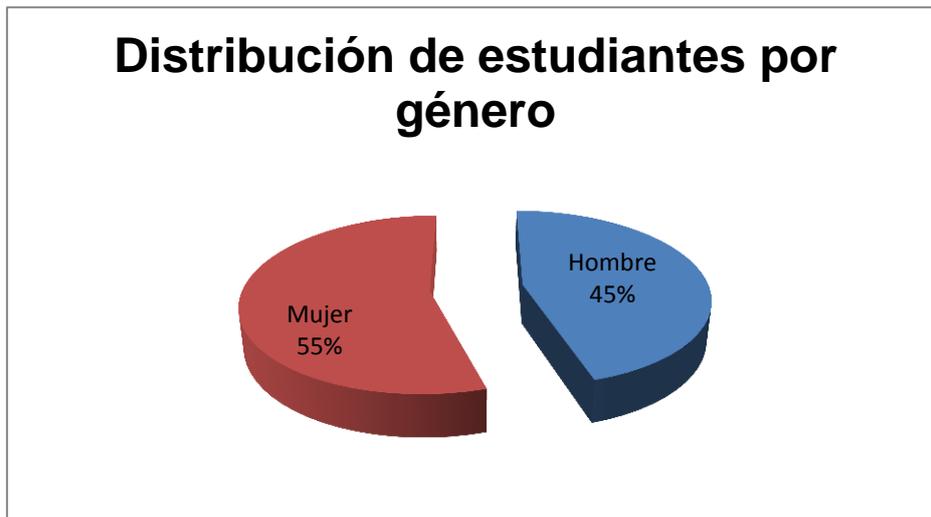


Figura 8: Distribución de estudiantes por género.

Fuente: Elaboración: Propia

La Universidad Técnica Particular de Loja oferta 26 titulaciones comprendidas en cuatro áreas: administrativa, biológica y biomédica, socio humanística y técnica, en las cuales la mayor proporción de estudiantes (26,8%) encuestados corresponde a la titulación de contabilidad y auditoría; mientras que la menor proporción (0,2%) corresponde a la titulación de ingeniería química. Ver Figura 9.

Distribución de estudiantes por titulación académica

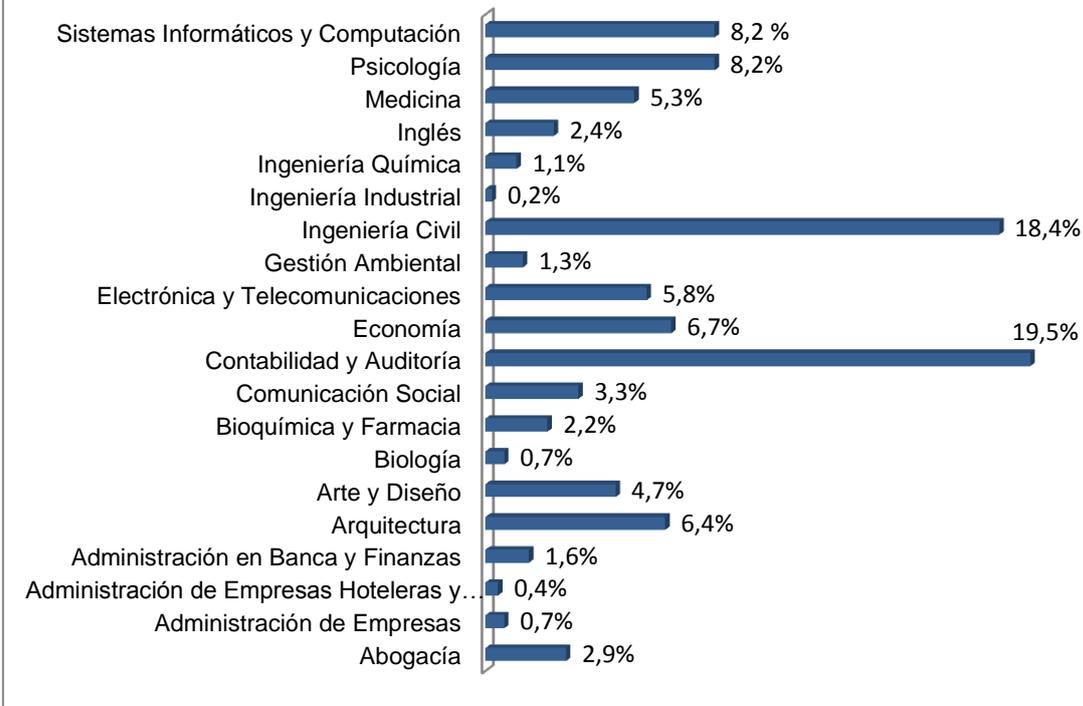


Figura 9: Distribución de estudiantes por titulación académica.
Fuente: Elaboración propia

Las edades estaban comprendidas entre 18 y 32 años ($M=21,23$; $DT=2,17$), con una gran concentración de datos por arriba de los 21 años. Ver Figura 10.

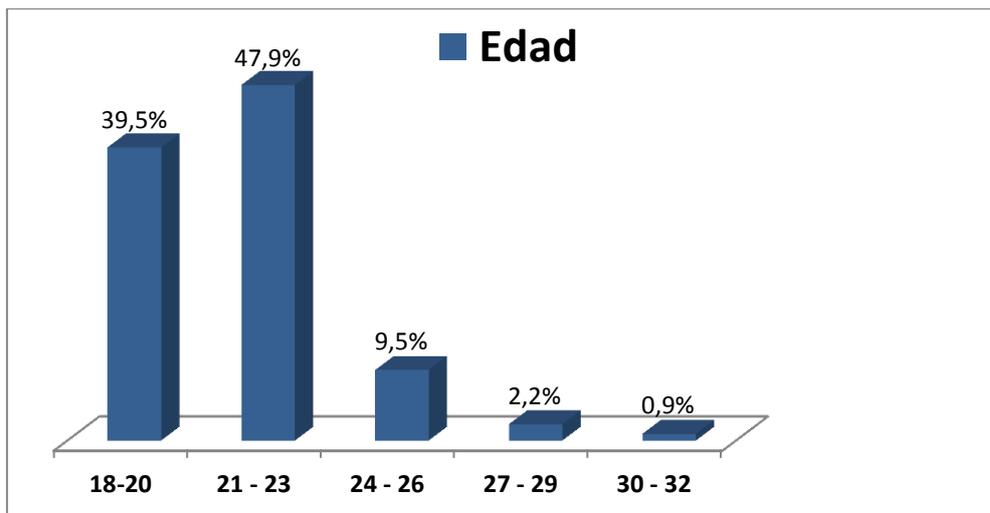


Figura 10: Distribución de estudiantes por edad
Fuente: Elaboración propia

Los niveles de ingreso de los estudiantes, se presentan en la Figura 11, se puede observar que la mayor proporción, (26,8%) corresponde a quienes tienen ingresos hasta 1000 dólares; la menor proporción (15,1%) corresponde a las familias que perciben más de 1500.

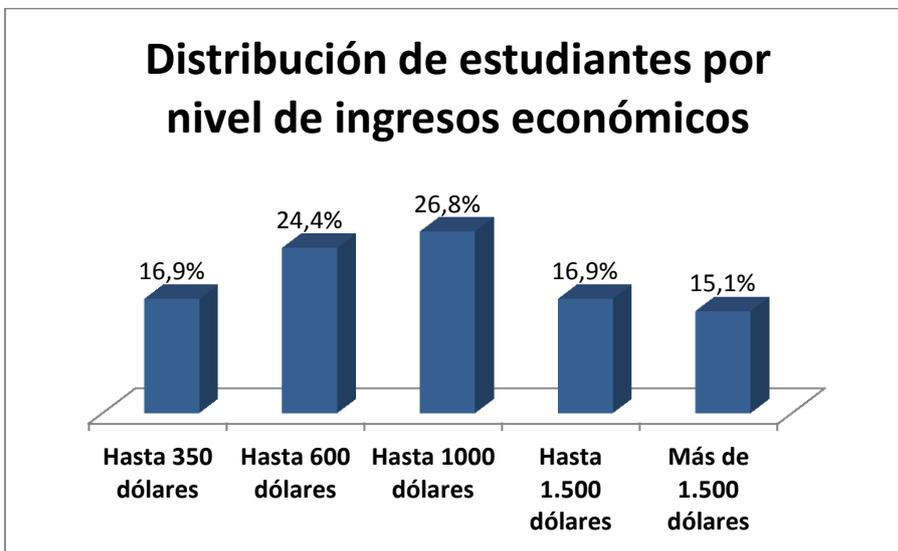


Figura 11: Distribución de estudiantes por nivel de ingresos económicos.
Fuente: Elaboración propia

4.2 Conocimiento y preferencias del estudiante

A continuación se presenta varios aspectos referentes a los conocimientos del manejo de internet y a las preferencias generales que tienen los estudiantes, en cuanto a lugar y tiempos de conexión, relacionándose a las posibilidades socioeconómicas que tiene el estudiante Utepelino.

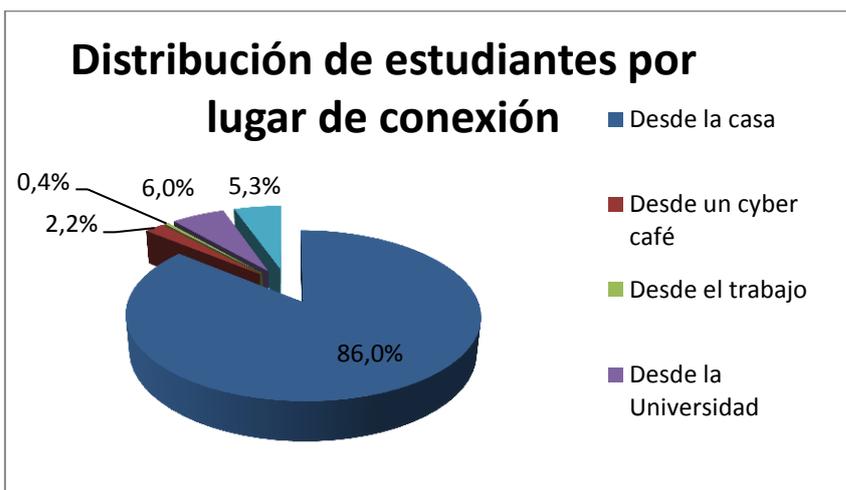


Figura 12: Distribución de estudiantes por lugar de conexión.
Fuente: Elaboración: Propia

Se puede observar (ver Figura 12) que el mayor porcentaje de estudiantes, se conecta desde la casa (86%) y el menor desde el trabajo (0,1%). Existe poco uso del internet en las instalaciones universitarias, también se nota que el uso del cyber café es escaso.

Los días de conexión a la red de los estudiantes se presentan en la Tabla 23, en ella se muestra que la mayoría, se conecta los 7 días a la semana (71,6%) mientras que el menor porcentaje de encuestados lo hace 2 días (0,9).

Tabla 23: Cantidad de días a la semana que se conectan a Internet los estudiantes

Días de conexión	Porcentaje de estudiantes
2	0,9
3	2,2
4	3,8
5	12,2
6	9,3
7	71,6
Total	100

Fuente: Elaboración: Propia

El nivel de conocimiento en el manejo de Internet del estudiante, oscila entre 1 que indica que éste es bajo y 10 que muestra que es alto. Se puede observar que el 26, 2% de estudiantes tiene un conocimiento en el nivel 8, mientras que el 3,8% no supera los niveles que se encuentran entre el 1 y 4. Ver Tabla 24.

Tabla 24: Distribución de estudiantes por el nivel de conocimiento en Internet.

Nivel de conocimiento	Porcentaje de estudiantes
1	0,2
3	0,9
4	2,7
5	6,2
6	9,5
7	20
8	26,2
9	17,5
10	16,9
Total	100

Fuente: Elaboración: Propia

La Figura 13 muestra que un 47,0% de estudiantes se conectan entre 4 a 6 horas diarias, y la menor parte se conecta entre 13 y 15 horas (0,2%), indicando que los encuestados destinan un considerable número de horas para el uso de internet.

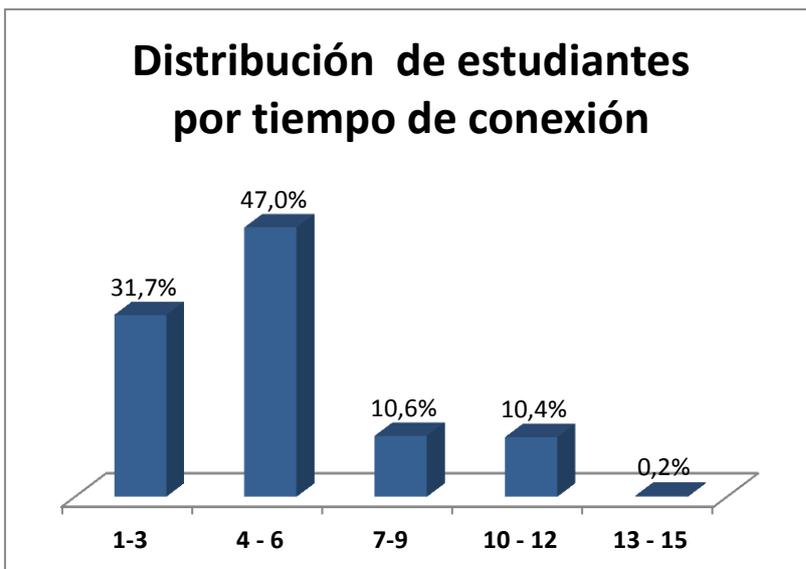


Figura 13: Distribución de estudiantes por tiempo de conexión.
Fuente: Elaboración: Propia

La mayor parte de estudiantes ha empezado a conectarse desde hace 5 años (29,3%), aproximadamente, un 22,2% tiene menos de 4 años de utilizar la red, el restante 48,5% de estudiantes tienen más de 5 años de experiencia en el uso de internet, con una tendencia a variar por debajo o por encima de 2,11. Ver Tabla 25.

Tabla 25: Años de uso de internet del estudiante.

Años de uso de internet	Porcentaje de estudiantes.
1	0,4
2	2,7
3	6,9
4	12,2
5	29,3
6	13,7
7	11,1
8	10,2
9	2,4
10	11,1
Total	100

Fuente: Elaboración Propia

4.3 Uso del internet en el ámbito académico y no académico

4.3.1 Usos académicos

En esta sección se muestra los resultados acerca del uso del internet en las actividades académicas, en lo referente a las asignaturas que se encuentran matriculados los estudiantes.

El promedio de ingreso a la plataforma virtual es de seis horas, con una desviación estándar de 2,89 horas, se evidencia que los estudiantes revisan un considerable número la información referente a sus estudios. Para evaluar el ingreso máximo se considera como máximo 14 horas y mínimo de 1 hora de ingreso. Ver Tabla 26.

Tabla 26: Ingreso a la plataforma virtual

Ingreso plataforma Virtual	Porcentaje de estudiantes
0,00 - 2,00	13,3
3,00 - 5,00	40,4
6,00 - 8,00	31
9,00 - 11,00	11,3
12,00 - 14,00	4
Total	100

Fuente: Elaboración: Propia

En la Tabla 27, podemos observar que en promedio los estudiantes realizan alrededor de 5 consultas al profesor a través de la plataforma virtual al mes. Se presenta una desviación estándar de 5,69 observando una dispersión más alta entre los datos. Poco uso de la consultas a través de la plataforma virtual.

Tabla 27: Consultas al profesor

Consultas al Profesor	Porcentaje de estudiantes
0,00 - 3,00	55,2
4,00 - 7,00	24,8
8,00 - 11,00	10,6
12,00 - 15,00	3,3
16,00 - 19,00	1,1
20,00 - 23,00	2,9
24,00 - 27,00	0,4
28,00 - 31,00	1,6
Total	100

Fuente: Elaboración Propia

En promedio los estudiantes realizan alrededor de 9 consultas a sus compañeros al mes a través de la plataforma virtual, el mayor número de consultas que efectúan los estudiantes son de 30. Podemos darnos cuenta que los estudiantes realizan más consultas a los compañeros (9 consultas) que a sus profesores (5 consultas) por mes.

En la Tabla 28 podemos observar la intensidad de descargas al mes de recursos educativos para las actividades académicas. El 76 (16,9%) descargan alrededor de 10 recursos educativos mientras que el 10,4% no realizan ninguna descarga de estos recursos. En promedio los estudiantes descargan alrededor de 6,65 recursos educativos que son apoyo importante para culminar cada ciclo académico

Tabla 28: Descargas de recursos educativos

Recursos educativos	Porcentaje de estudiantes
0,00 - 2,00	26,6
3,00 - 5,00	28,2
6,00 - 8,00	12,6
9,00 - 11,00	17,7
12,00 - 14,00	1,6
15,00 - 17,00	6,2
18,00 - 20,00	7,1
Total	100

Fuente: Elaboración: Propia

La mayor parte de los estudiantes realizan 2 foros al mes, observando que la participación dentro de los mismos es baja, existe el 16,9% de estudiantes que no lo realiza. Ver Figura 14.

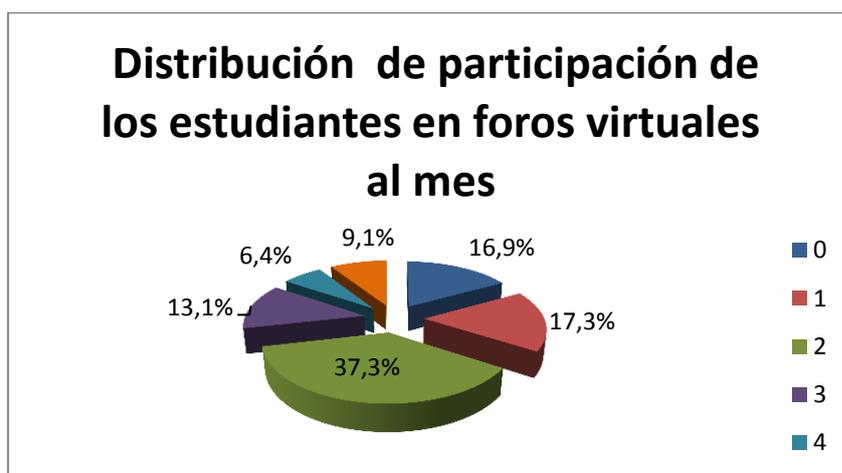


Figura 14: Distribución de participación de los estudiantes en foros virtuales al mes

Fuente: Elaboración Propia

Así mismo, los resultados indican que los estudiantes miran alrededor de 3 a 5 videos académicos en la plataforma de YouTube (39,0%) al mes, en tanto que, el 1,1 % observan alrededor de 14 videos. Ver Figura 15.



Figura 15: Videos académicos que el estudiante mira en YouTube al mes.
Fuente: Elaboración: Propia

En los estudiantes encuestados, la participación promedio es de 3 tweets académicos al mes, se puede observar que la mayor parte, no utiliza esta herramienta o participa de 1 a 2 veces (66,3%) en esta actividad. Ver Figura 16.

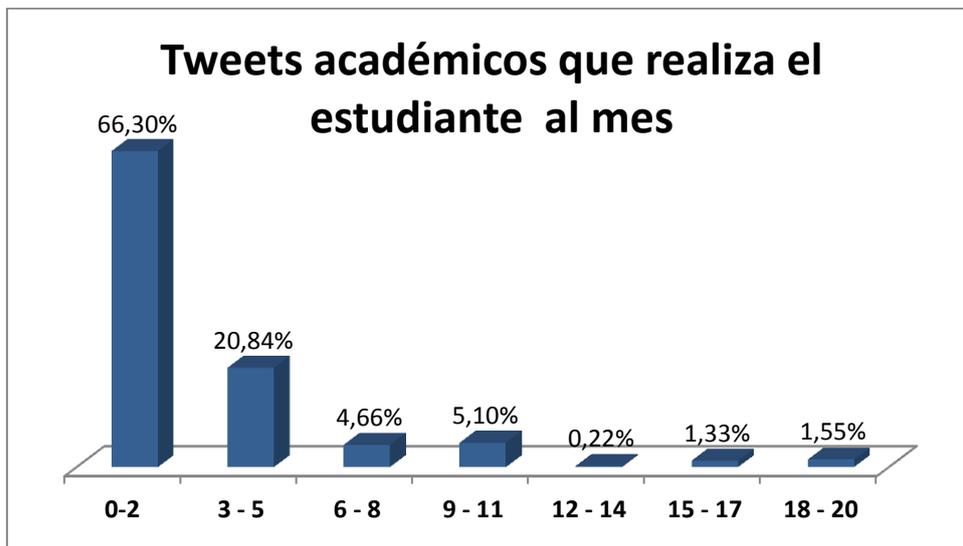


Figura 16: Tweets académicos que realiza el estudiante al mes.
Fuente: Elaboración: Propia

Con respecto a la actividad del uso de chat para fines académicos, se puede notar que el promedio de su utilización es de 7,59 por mes, con una tendencia a variar por encima o

debajo de 7 chats, Existe una variabilidad en los datos de 32,047. Se observa además que la moda es de 10 chats (28,6%). Ver Tabla 29.

Tabla 29: Uso del chat académicos por horas al mes

Chats académicos	Porcentaje de estudiantes
0,00 - 2,00	24,2
3,00 - 5,00	20,2
6,00 - 8,00	9,8
9,00 - 11,00	29,3
12,00 - 14,00	2,2
15,00 - 17,00	5,3
18,00 - 20,00	9,1
Total	100

Fuente: Elaboración: Propia

En la Figura 17 se puede observar que el 31% de los estudiantes utilizan de 20 a 29 horas el internet, para búsqueda de información académica al mes, mientras que el 3% realizan más de 60 búsquedas aproximadamente. En promedio realizan 19 indagaciones académicas, con una tendencia a variar por encima o por debajo de 13,81.



Figura 17: Información académica que buscan los estudiantes en internet.

Fuente: Elaboración: Propia

Con respecto a si los estudiantes utilizan la biblioteca virtual, es decir, utilizan los recursos que esta ofrece, como acceso a revistas científicas o libros online, podemos observar que hay poco uso de esta herramienta, el 28,4% no accede a la biblioteca virtual, el promedio de horas al mes es de 5,36 con una tendencia a variar superior o inferior a 7,16. Ver Figura 18.

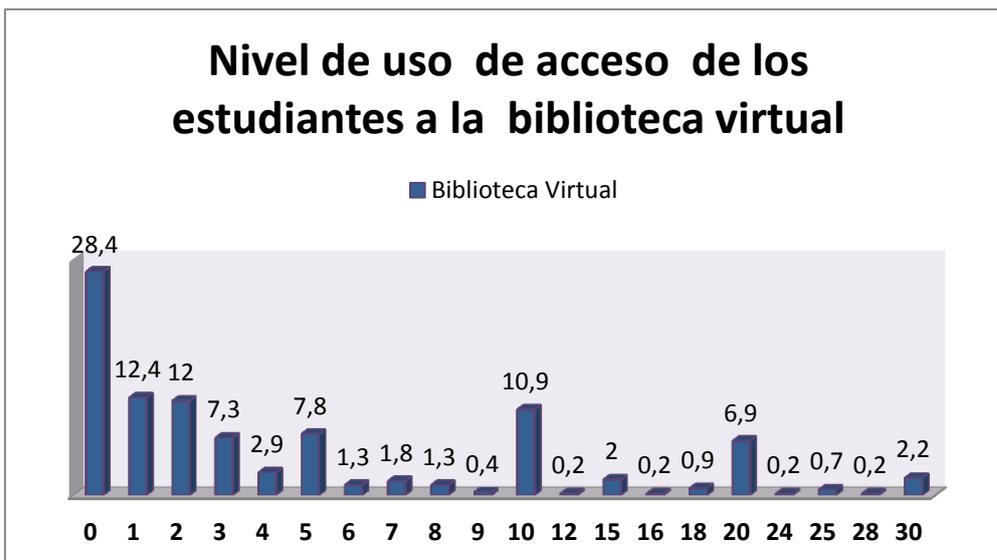


Figura 18: Nivel de uso de acceso de los estudiantes a la biblioteca virtual.
Fuente: Elaboración: Propia

4.3.2 Usos no académicos

Continuando con la investigación, seguidamente se exponen los resultados referentes al uso del internet en el ámbito de entretenimiento y diversión a los que se dedican los estudiantes, en la Tabla 30 se puede observar que el promedio de chats por diversión es de 9 horas semanales.

Tabla 30: Uso del chat por diversión por horas a la semana

Chat de diversión	Porcentaje de estudiantes
0,00 - 3,00	19,5
4,00 - 7,00	24,8
8,00 - 11,00	30,8
12,00 - 15,00	10
16,00 - 19,00	0,4
20,00 - 23,00	8,9
24,00 - 27,00	1,3
28,00 - 31,00	4,2
Total	100

Fuente: Elaboración Propia

Se determinó que el 23,9% de los estudiantes utilizan las redes sociales alrededor de 10 horas por semana, el promedio es de utilización es de 11,67 con una desviación estándar de 8,88 horas. El 1.5% no utilizan las redes sociales para el entretenimiento y diversión como podemos ver en la Tabla 31.

Tabla 31: Uso de redes sociales por horas a la semana

Redes Sociales	Porcentaje de estudiantes.
0-5,00	28,2
6,00 - 11,00	37,9
12,00 - 17,00	12,6
18,00 - 23,00	11,3
24,00 - 29,00	2
30,00 - 35,00	4,7
36,00 - 41,00	3,3
Total	100

Fuente: Elaboración propia

Los estudiantes utilizan en promedio 5 horas para juegos en línea a la semana, manteniendo una tendencia a variar superior o inferior a 7 horas, también se observa que el 54,3 % de la población estudiantes no utiliza los juegos en línea, mientras que tan solo el 11% de estudiantes utilizan más de 20 horas esta diversión online.

Las horas que descarga música, videos y programas a la semana el estudiante es en promedio de 4 horas, la tendencia de variación es superior o inferior a 4. Como se observa la Tabla 32, los estudiantes que descargan entre 4 y 7 horas semanales se encuentran en el 21,3%.

Tabla 32: Uso de herramientas para descargar música, videos y programas

Descarga música, videos, programas	Porcentaje de estudiantes
0,00 - 3,00	59,9
4,00 - 7,00	81,2
8,00 - 11,00	97,4
12,00 - 15,00	100

Fuente: Elaboración: Propia

El promedio de horas usadas para observar videos para entretenimiento en la cuenta YouTube está alrededor de los 8 videos semanales, con una desviación estándar de alrededor de 7 videos, tal como lo indica la Tabla 33.

Tabla 33: Uso de la cuenta YouTube para entretenimiento

Videos entretenimiento YouTube	Porcentaje de estudiantes
0,00 - 3,00	31,3
4,00 - 7,00	31,7
8,00 - 11,00	20,4
12,00 - 15,00	4,2
16,00 - 19,00	0,4
20,00 - 23,00	7,1
24,00 - 27,00	1,6
28,00 - 31,00	3,3
Total	100

Fuente: Elaboración: Propia

4.4 Uso de redes sociales

En la encuesta planteada, se pregunta a los estudiantes el número de amigos o seguidores que tiene en algunas redes sociales obteniendo los siguientes resultados:

El 98,9% de estudiantes tienen menos de 1000 seguidores twitter, el 74,3% de los estudiantes tiene menos de 1000 amigos o contactos en la red social facebook, y el 98,2% de los estudiantes tiene menos de 1000 contactos en linkelink. Ver Figura 19.

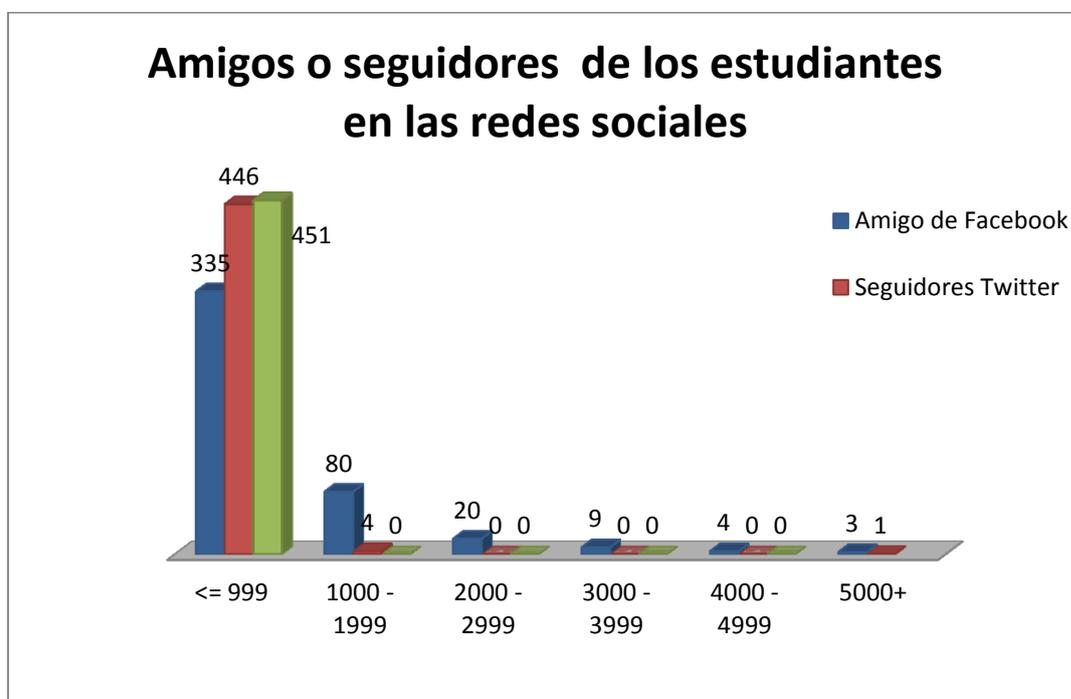


Figura 19: Amigos o seguidores de los estudiantes en las redes sociales

Fuente: Elaboración: Propia

4.5 Rendimiento Académico

Para el rendimiento académico se consideró la pregunta 17 del cuestionario, que se basa en el número de asignaturas aprobadas y las que el alumno reprobó, para construir la variable (rendimiento_academico) se la realiza restando el número de asignaturas matriculadas del número de asignaturas aprobadas, (Torres Díaz, 2012), posteriormente se procede a categorizar la variable en dos niveles (rendimiento_acad), considerando el nivel 1 representa a los estudiantes aprobados, y el nivel 2 representa a los estudiantes que han reprobado. Ver Figura 20.

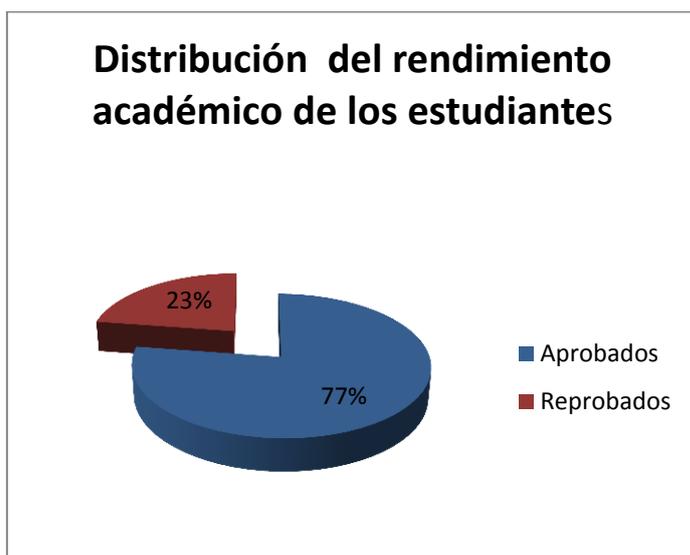


Figura 20: Distribución del rendimiento académico de los estudiantes
Fuente: Elaboración: Propia

El rendimiento académico presenta una asociación significativa, con la variable eda (Tau C=0,147, $p=0,001$; $p<0.05$). Se determina una existencia de asociación significativa entre el género y el rendimiento académico (Tau B=0,046, $p=0,000$; $p<0.05$) El rendimiento académico y el nivel de ingreso no presentan una asociación significativa (Tau C=-0,028, $p=0,511$; $p<0.05$)

4.6 Relaciones entre variables

Se encontró una relación significativa entre edad y los años que se conectan los estudiantes a internet (R de Pearson =0,196, $p<0,00$; $p<0.05$). La dependencia de la variable eda sobre la variable de año_exp alcanza un 3,84%. Así mismo se determinó que existe una relación entre la edad y las horas que el estudiante chatea sobre temas académicos (R de Pearson =-0,116, $p<0,01$; $p<0.05$) esto alcanza un coeficiente de determinación del 1,34%.

Se determinó la existe de una relación significativa entre la edad con las asignaturas que el estudiante se matriculó en el semestre anterior del (R de Pearson =-0,254, $p<0,00$; $p<0.05$) y las que reprobó (R de Pearson =-0,299, $p<0,000$; $p<0.05$) La dependencia de la variable eda

sobre la variable cic_asi_mat alcanza un coeficiente de determinación es 6,45 % y de la variable cic_asi_apr el coeficiente de determinación es de 8,94 %.

Con respecto a los ingresos y al lugar de conexión habitual a de internet de los estudiantes, presentan una asociación significativa ($\chi^2 = 30,21$, $p < 0,017$; $p < 0,05$). Existe una relación entre la variable ingresos y los días que el estudiante se conecta a la semana (R de Pearson = 0,226, $p < 0,000$; $p < 0,05$), esto alcanza un coeficiente de determinación del 5,10 %. Asimismo se encontró una relación significativa entre la variable ingresos y el nivel de conocimiento en el manejo del internet (R de Pearson = 0,228, $p < 0,000$; $p < 0,05$). La dependencia de la variable ingresos alcanza un coeficiente de determinación es 5,1984 % sobre la variable niv_con.

4.7 Perfiles de estudiantes

4.7.1 Perfil de estudiantes en función a las actividades académicas

Para realizar la disminución de variables con lo referente a las actividades académicas del estudiante, se realiza un análisis previo de los datos, es decir, se toman en cuenta las variables que más destacan, ya que son los factores que más influyen en el proceso de aprendizaje del estudiante. Ver Figura 21. Las actividades agrupadas son las siguientes:

- Busca información académica.
- Chat sobre temas académicos.
- Ingreso a la plataforma virtual.
- Videos académico YouTube.
- Descargas de la plataforma virtual.
- Consultas a los compañeros.

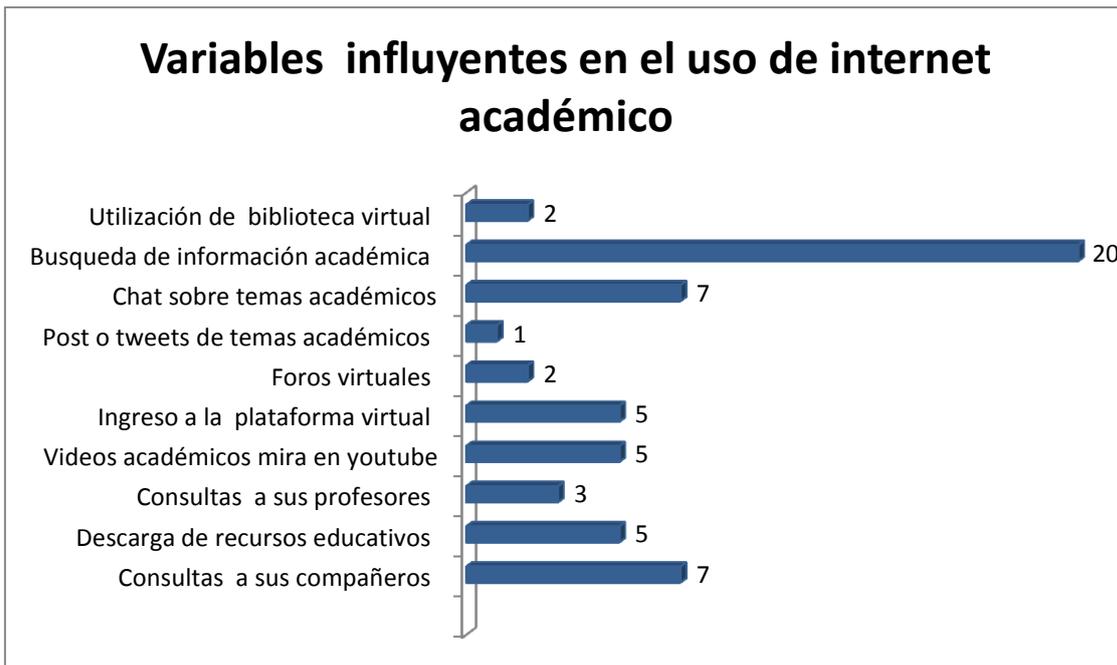


Figura 21: Variables influyentes en el uso de internet académico
Fuente: Elaboración: Propia

Consecutivamente se procedió a realizar la clasificación de los estudiantes, se recurrió al método de análisis cluster k-medias; para determinar los grupos se utilizó el programa SPSS versión 17. De los 451 registros ingresados de los estudiantes se procedió a realizar los cluster con base en las variables reducidas. Se realizó la clasificación con 2,3, 4 y 5 cluster Ver Anexo 2, para garantizar una correcta agrupación.

Luego de realizar la clasificación y agrupar a los estudiantes, se procedió a realizar la verificación de los cluster mediante el análisis discriminante, en la Tabla 34 se puede observar el porcentaje de exactitud, seguidamente se escogió el grupo 2, ya que la clasificación es la más precisa. Ver Anexo 3.

Tabla 34: Porcentajes de exactitud de la clasificación-Académica

Grupos	Porcentaje de exactitud
2	99,6%
3	97,10%
4	98,90%
5	97,30%

Fuente: Elaboración Propia

La clasificación de los grupos se la puede observar en la Figura 22.

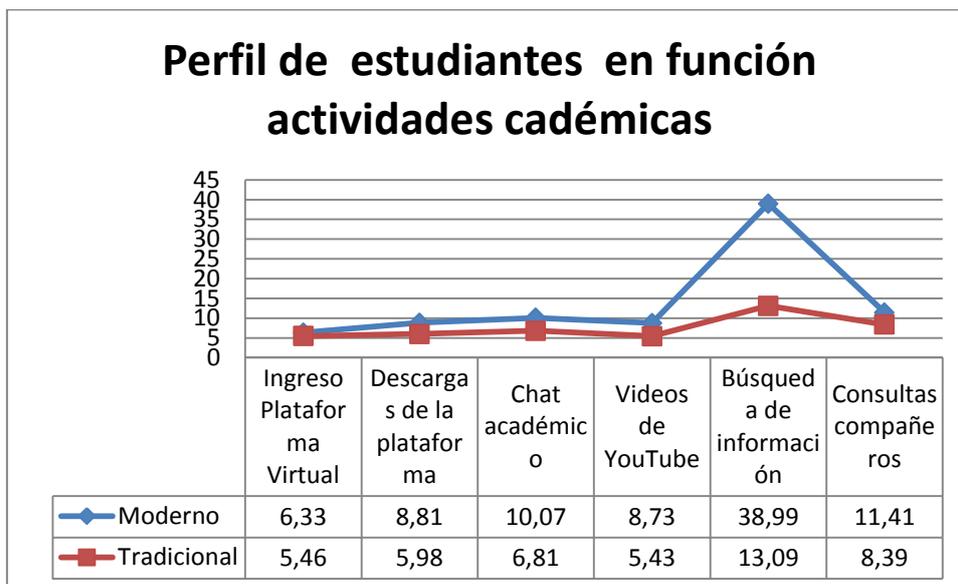


Figura 22: Perfiles de estudiantes en función de las actividades académicas

Fuente: Elaboración: Propia

El primer grupo denominado moderno está conformado por el 23,9% del total analizado, cuya característica es presentar valores mayores al grupo tradicional. Los estudiantes de este grupo tienen mayor uso de búsqueda de información académica, chat sobre temas académicos y mayor uso de horas en la visualización de videos académicos. Así mismo estos estudiantes realizan más consultas a compañeros. La edad promedio de este grupo es de 21,41 años.

El grupo 2 incluye al 76,1% de estudiantes y utiliza muy poco los recursos de internet como medio de aprendizaje, es por esta razón que lo denominamos tradicional, y se caracteriza por tener menos valores en la descarga de recursos educativos, consultas académicas bajas, el escaso ingreso a la plataforma virtual, chat sobre temas académicos bajos y la exigua búsqueda de información académica en internet. La edad promedio de este grupo es de 21,18.

4.7.2 Perfil de estudiantes en función a las actividades de entretenimiento

Son cinco las preguntas que se realizaron a los estudiantes, con respecto al uso de internet en las actividades académicas y de entretenimiento. De estas variables se procedió a realizar un análisis previo, para ver las más relevantes con respecto a las variables que más destacan, ya que son los factores que más influyen en el proceso de entretenimiento y diversión de los estudiantes. Ver Figura 23, quedando las siguientes variables:

- Chat por diversión
- Utilización de redes sociales
- Videos que el estudiante ve el YouTube para su entretenimiento.

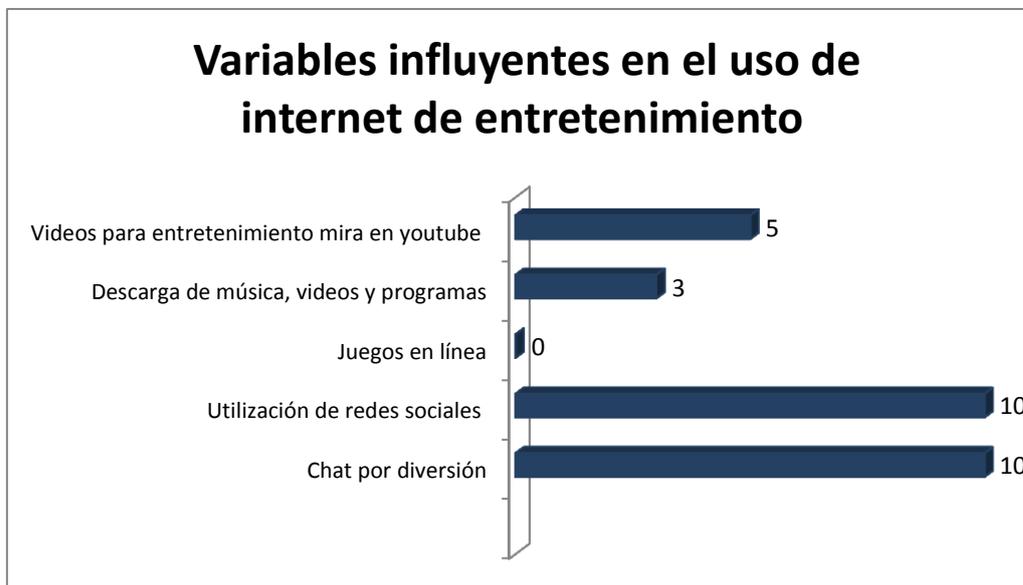


Figura 23: Variables influyentes en el uso de internet de entretenimiento.
Fuente: Elaboración Propia

Para clasificar los grupos de los estudiantes se utilizó el método de análisis de k-medias, mencionado anteriormente, se procedió a realizar clasificaciones con 2, 3, 4 y 5 grupos, las variables a clasificar fueron las de chat por diversión, utilización de redes sociales y videos que ve el estudiante en YouTube para entretenimiento y diversión. Ver Anexo 4.

Luego de clasificar se procedió a realizar la verificación utilizando la técnica del análisis discriminante. Como se puede observar en la Tabla 35 el porcentaje más exacto es el del cluster del grupo 2, por esta razón, se trabajó con esta clasificación. Ver Anexo 5.

Tabla 35: Porcentajes de exactitud de la clasificación-Entretenimiento

Grupos	Porcentaje de exactitud
2	99,6%
3	98,90%
4	97,30%
5	98,20%

Fuente: Elaboración Propia

A continuación se presenta la descripción de cada grupo, en cuanto a sus características más importantes. Ver Figura 24.

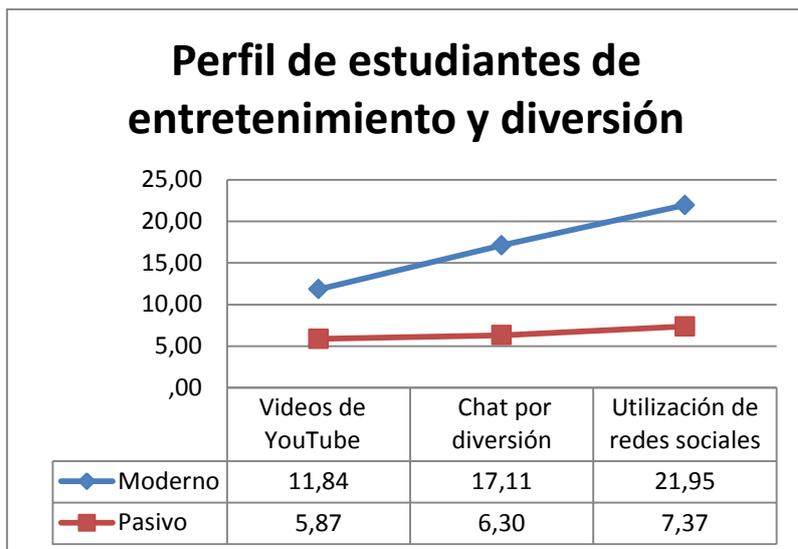


Figura 24: Perfiles de estudiantes en función de las actividades de entretenimiento
Fuente: Elaboración: Propia

La inclusión de horas semanales que el estudiante utiliza en chatear por diversión, las que usa en redes sociales y la totalidad de videos semanales que observa para entretenimiento clasifica a los estudiantes en dos grupos.

El grupo “moderno” incluye al 29,5% de estudiantes, su principal características es el mayor uso de las redes sociales y chat, por diversión utilizan más de 17 horas semanales en chatear y además observar más de 11 videos semanales. La edad promedio de este grupo es de 21 años.

El grupo “pasivos” constituye el 70,5% de estudiantes, este grupo utiliza muy poco los recursos de internet, como medio de entretenimiento y diversión. Sus características son la poca utilización de visualización de videos en la red social YouTube, y el chat por diversión, utilizan aproximadamente 7,37 horas en el uso de redes sociales. La edad promedio del grupo pasivo es de 21 años.

4.7.3 Perfil de estudiantes en función al nivel, uso de dispositivos

En una sola pregunta se comprende siete ítems, se procedió a realizar un análisis previo y se toman las variables que influyen más en los estudiantes, con la ayuda de la mediana. Ver Figura 25. Extrayendo las más influyentes quedaron:

- Smartphone con cámara fotográfica y acceso a internet .
- Computador portátil.
- Teléfono móvil con acceso a internet.

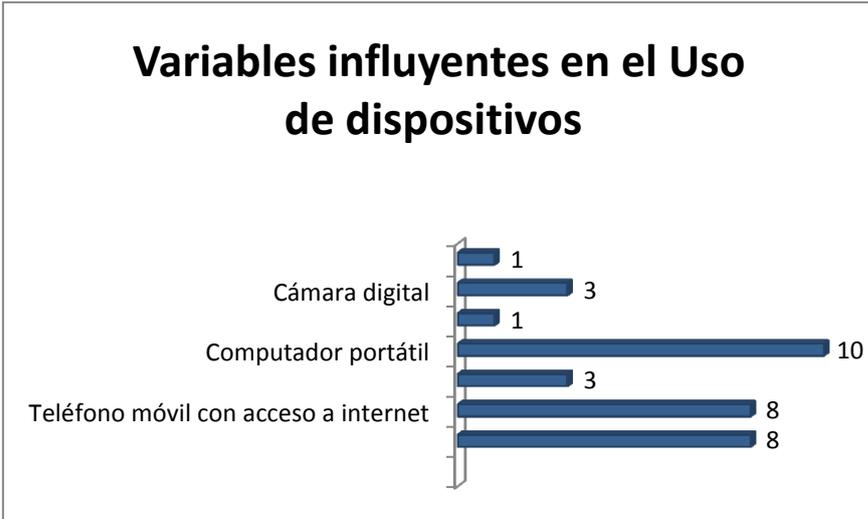


Figura 25: Variables influyentes en el Uso de dispositivos
Fuente: Elaboración: Propia

Para la clasificación de los estudiantes se recurre al método de análisis cluster k-medias; para determinar los grupos, se utilizó el programa SPSS versión 17. Se realizó la clasificación con 2, 3 y 4 cluster (Anexo 6) para garantizar una correcta agrupación.

Luego de realizar la clasificación y agruparlos mediante perfiles a los estudiantes se procedió a realizar la verificación de los cluster mediante el análisis discriminante, en la Tabla 36 se observa el porcentaje de exactitud, para el propósito se escogió el grupo dos. Ver Anexo 7.

Tabla 36: Porcentajes de exactitud de la clasificación-Nivel uso de dispositivos

Grupos	Porcentaje de exactitud
2	97,30%
3	96,50%
4	96,70%

Fuente: Elaboración Propia

La agrupación de dos grupos se muestra a continuación. Ver Figura 26.

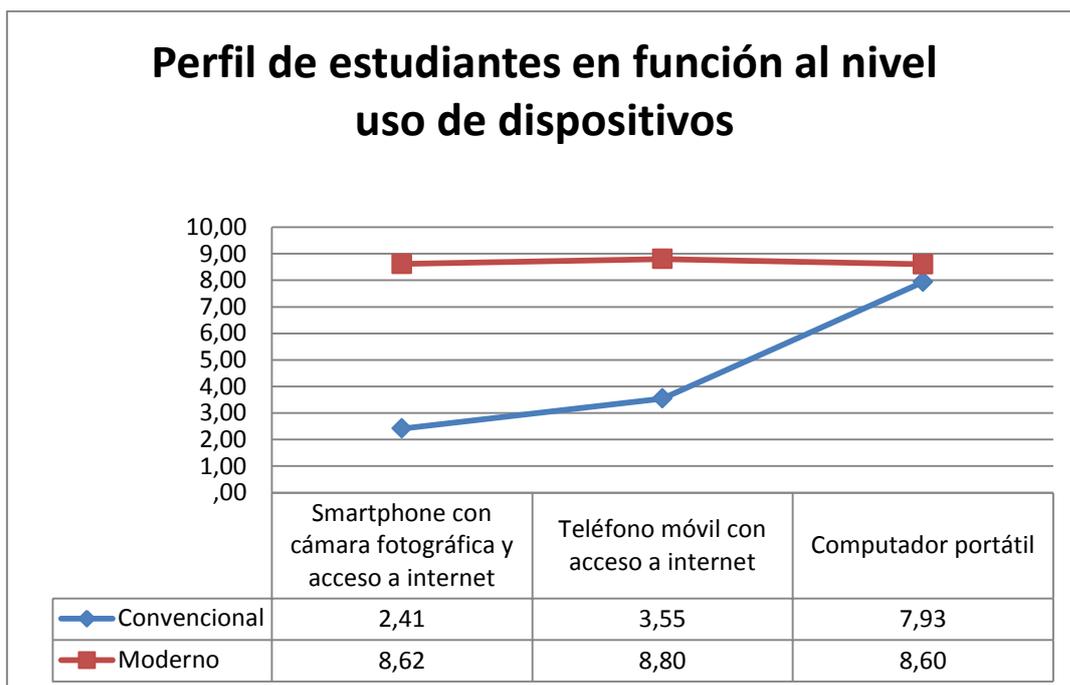


Figura 26: Perfiles de estudiantes en función al nivel de uso de dispositivos.

Fuente: Elaboración: Propia

El uso de dispositivos tecnológicos clasifica a los estudiantes en dos grupos:

El grupo 1 o de bajo uso está conformado por el 40,1% de estudiantes que utilizan en menor parte estos dispositivos electrónicos, por esta razón, lo denominaremos como grupo tradicional, y se caracteriza por utilizar en menor parte los dispositivos electrónicos. La edad promedio es de 21,49.

El grupo dos, lo conforman el 59,9% de estudiantes tienen un mejor acceso a los dispositivos, se caracteriza por tener elevado acceso al uso de los dispositivos. Por esta razón, lo hemos considerado como grupo moderno. La edad promedio de este grupo es de 21,06 similar a la del grupo convencional.

4.8 Comprobación de las hipótesis

4.8.1 Perfil académico y perfil de entretenimiento

Perfil Académico

En este apartado se comprobó la hipótesis uno, planteada en esta investigación, que sostiene que el nivel de ingresos determina como se utiliza internet para el aprendizaje.

Los perfiles académicos construidos anteriormente, diferencian a los estudiantes con base en el uso que le otorgan a la tecnología, en actividades relacionadas con su formación universitaria, estos perfiles son: moderno y tradicional.

Para establecer relación entre la variable dependiente y la independiente a través de las pruebas Chi Cuadrado, los resultados presentan un valor significativo ($\chi^2=15,999$, $p=0,042$; $p<0,05$), por lo tanto, se concluyó que existe relación significativa ($P <0,05$), entre estas variables. Ver Anexo 8.

La comprobación del modelo se lo realiza a través de la regresión logística binaria (Anexo 10), ya que la variable dependiente es una variable dummy, es decir asume dos categorías. En primera instancia la codificación de la variable dependiente, resultan así moderno: 0 tradicional: 1, La estructura de la variable muestra que el 76,1% (343) pertenecen al perfil de tradicionales y el restante el 23,9% (108) al perfil de estudiantes modernos.

No todas las variables que no se encuentran en la ecuación resultan no significativas, en el nivel 1(hasta 350 dólares) $p=0,029$ y el nivel 3(hasta 1.000 dólares) de la variable independiente resulta significativa $p=0,016$, es decir, si mejora la probabilidad de aciertos. Es por esta razón, que seguiremos con el modelo.

Para la comprobación del modelo se utilizó la prueba de Ómnibus, dando una significación estadística ($X^2= 10, 637$, $p=0, 032$; $p<0,05$), esto nos indica que hay una mejora significativa en la predicción de probabilidad de ocurrencia de las categorías de la variable ingresos. El estadístico de Nagelkerke explica la varianza representado en este modelo que es de 0,035 (3,5%).

La prueba de Hosmer y Lemeshow, verifica la Hipótesis nula de continuación:

- H_0 : Valores observados son iguales a los valores predichos
- H_1 : Valores observados son diferentes de los valores predichos

El valor de significancia es de ($X^2=0,000$, $p=1,000$), el valor de $p=1,000$ indica que se acepta H_0 , esto implica que el modelo se ajusta a los datos.

Para la verificación de los coeficientes de la variable independiente se utilizó la prueba de Wald, (Ver Tabla 37) no todas las variables alcanzan un valor significativo ($p<0,05$), sin embargo en dos categorías si contribuyen en que los estudiantes se encuentren en el grupo tradicional o moderno. Por tal caso se rechaza la hipótesis nula y se deduce que el nivel de ingresos si determina como se utiliza internet para el aprendizaje.

Tabla 37: Coeficientes del modelo de regresión de perfiles de uso de Internet en actividades académicas e Ingresos

		Variables en la ecuación					I.C. 95% para EXP(B)		
		B	E.T.	Wald	Gl	Sig.	Exp(B)	Inferior	Superior
Paso 1 ^a	Ing			10,393	4	,034			
	ing(1)	,628	,369	2,896	1	,089	1,873	,909	3,860
	ing(2)	1,157	,364	10,079	1	,001	3,181	1,557	6,497
	ing(3)	,658	,331	3,947	1	,047	1,931	1,009	3,696
	ing(4)	,487	,362	1,811	1	,178	1,628	,801	3,310
	Constante	,542	,252	4,650	1	,031	1,720		

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

Fuente: Elaboración Propia

El modelo encontrado, se puede expresar de la siguiente forma:

$$Pr(Uso_Internet_academico) = \frac{1}{1 + e^{-(0,542 - 0,628Ingresos(1) - 1,157Ingresos(2) - 0,658Ingresos(3) - 0,487Ingresos(4))}}$$

La puntuación de Wald para el modelo probado indica que la variable uso_internet_academico aporta significativamente a la predicción de la variable ingreso, los resultados obtenidos se pueden generalizar a la población.

El odd ratio (OR) de pertenecer al perfil de tradicional con respecto al perfil moderno es de 3,181 veces mayor cuando el estudiante pertenece al nivel de ingresos 3, es decir, de hasta 1000 dólares, respecto al nivel 1 (OR=3,181, (IC 95% 1,557-3,860), p=0,01); es de 1,931 veces mayor si el estudiante pertenece al nivel de ingreso 4 es decir de hasta 1.500 dólares, respecto al 1 (OR=1,931, (IC 95% 1,009-3,696), p= 0,047).

Con los valores encontrados se puede señalar que los estudiantes con mayores ingresos tienen más posibilidades de utilizar el internet para actividades académicas.

Perfil de Entretenimiento

En este componente se contrasta la hipótesis 2, que sustenta que el nivel de ingresos determina como se utiliza internet para entretenimiento.

Se utiliza la técnica estadística chi-cuadrado de Pearson, dando como resultado ($\chi^2=9,522$, p=0,300) con este valor de significancia (P <0,05), se puede deducir que el uso de internet para entretenimiento no depende de los ingresos. Ver Anexo 8.

La codificación de la variable dependiente es este caso, uso de internet para entretenimiento, tiene dos niveles que resultan así 0: Modernos y 1: Pasivos. La estructura de la variable muestra que el 29,5 % (133) son estudiantes Modernos, mientras que el 70,5% (318) pertenecen al grupo de estudiantes Pasivos. Ver Anexo 11.

Los resultados de la prueba ómnibus, indican que el uso de internet para entretenimiento no alcanzan un valor significativo ($X^2=4,401$, $p=0,355$; $p<0,05$). El estadístico R^2 Nagelkerke con valor de 0,014 (1,4%) indica que no fue aceptable el ajuste del modelo.

La prueba de Hosmer y Lemeshow, verifica la Hipótesis nula de continuación,

- H0: Valores observados son iguales a los valores predichos
- H1: Valores observados son diferentes de los valores predichos

Se observa que ($X^2=0,000$, $p=1,000$), el valor de $p=1,000$ indica que se acepta H0, es decir que valores observados son iguales a los encontrados, se considera el modelo.

Tabla 38: Coeficientes del modelo de regresión de perfiles de uso de Internet en actividades de entretenimiento e Ingresos

		Variables en la ecuación						I.C. 95% para EXP(B)	
		B	E.T.	Wald	gl	Sig.	Exp(B)	Inferior	Superior
Paso 1 ^a	Ing			4,453	4	,348			
	ing(1)	,487	,362	1,811	1	,178	1,628	,801	3,310
	ing(2)	,485	,332	2,135	1	,144	1,624	,847	3,112
	ing(3)	,481	,325	2,184	1	,139	1,617	,855	3,058
	ing(4)	,054	,347	,024	1	,877	1,055	,534	2,085
	Constante	,542	,252	4,650	1	,031	1,720		

Fuente: Elaboración propia

El modelo encontrado, se puede expresar de la siguiente forma:

$$\Pr(\text{Uso entretenimiento}) = \frac{1}{1 + e^{-(0,5420,487\text{Ing}(1)-0,485\text{Ing}(2)-0,481\text{Ing}(3)-0,054\text{Ing}(4))}}$$

Para la verificación de los coeficientes de la variable independiente se utilizó la prueba de Wald, (Ver Tabla 38) todas las variables alcanzan un valor no significativo ($p<0,05$), es decir que no contribuyen en que un estudiante se encuentre en el grupo moderno o pasivo.

Los coeficientes de la variable independiente son negativos y su Or radio menor a la unidad.

4.8.2 Usos de internet y rendimiento académico

Uso académico de internet y rendimiento académico.

La comprobación del modelo se realizó a través de las pruebas Chi Cuadrado, los resultados presentan un valor de Chi cuadrado no significativo en ninguno de los dos casos. Ver Anexo 9.

En la siguiente sección se contrasta adicionalmente la hipótesis 3, que establece que el uso de la tecnología en el aprendizaje incide en el rendimiento académico. Como se menciona anteriormente se utiliza regresión logística binaria por la variable dependiente (dicotómica).

La codificación de la variable dependiente es este caso rendimiento académico, tiene dos niveles que resultan así 0: aprobado y 1: reprobado. La estructura de la variable muestra que el 77,4% (349) son estudiantes aprobados mientras que el 22,6% (102) pertenecen al grupo de estudiantes reprobados. Ver Anexo 12.

Para este modelo, el análisis de regresión logística en el bloque 0 indica que hay un 77,4% de probabilidad de acierto en el resultado de la variable dependiente, si se asume que los estudiantes en su rendimiento académico aprueban las materias.

Los indicadores muestran que el rendimiento académico no incide con el uso de Internet académico, esto explica que la significancia del modelo es mayor a 0,05, es decir, el modelo no es significativo. ($R^2=0,171$, $p<0,680$; $p<0,05$). Esto se evidencia con el estadístico R^2 Nagelkerke, el cual nos da un valor de 0,010 que equivale al 1% de la varianza. Sin embargo, la bondad de ajuste del modelo a de la prueba de Hosmer y Lemeshow no da un valor de significancia esto implica que el modelo se ajusta a los datos.

Tabla 39: Coeficientes del modelo de regresión de rendimiento académico

		Variables en la ecuación					I.C. 95% para EXP(B)		
		B	E.T.	Wald	gl	Sig.	Exp(B)	Inferior	Superior
Paso 1 ^a	Cluster_acade_2(1)	,108	,260	,172	1	,678	1,114	,669	1,854
	Constante	-1,257	,130	93,404	1	,000	,285		

a. Variable(s) introducida(s) en el paso 1: Cluster_acade_2.

Fuente: Elaboración propia

El modelo encontrado, se puede expresar de la siguiente forma:

$$\Pr(\text{rendimiento}_{\text{academico}} = 1) = \frac{e^{1,76}}{1 + e^{-(-1,257 - 0,257 \text{clusaca2}(1))}}$$

En la Tabla 39 se observa que ninguno de los valores de las variable es significativo, los coeficientes de la variable independiente son, Clus_aca2 (1)= (OR=1,114, (IC 95% 0,669-1,854), p=0,678).

Uso de entretenimiento de internet y rendimiento académico.

Posteriormente se procedió a comprobar la a hipótesis 4 que sostiene que el uso de internet para entretenimiento, incide en el rendimiento académico, realizado en primera instancia el chi cuadrado resulta no significativa ($X^2=34,478$, $p=0,664$; $p<0.05$), Ver Anexo 9. Como se menciona anteriormente se utiliza regresión logística binaria por la variable dependiente (dicotómica). Ver Anexo 13

La codificación de la variable dependiente en este caso rendimiento académico, tiene dos niveles que resultan así 0: aprobado y 1: reprobado. La estructura de la variable muestra que el 77,4% (349), son estudiantes aprobados mientras que el 22,6% (102) pertenecen al grupo de estudiantes reprobados.

Para este modelo, el análisis de regresión logística en el bloque 0 indica que hay un 77,4% de probabilidad de acierto en el resultado de la variable dependiente, si se asume que los estudiantes en su rendimiento académico aprueban las materias.

Los indicadores muestran que el rendimiento académico, no incide con el uso de Internet académico, esto explica que la significancia del modelo es mayor a 0,05, es decir, el modelo no es significativo. ($R^2=2,085$, $p<0,149$; $p<0.05$). Esto se evidencia con el estadístico R^2 Nagelkerke el cual, nos da un valor de 0,07 que equivale al 7% de la varianza. Sin embargo, la bondad de ajuste del modelo a través de la prueba de Hosmer y Lemeshow no da un valor de significancia, esto implica que el modelo se ajusta a los datos.

Tabla 40: Coeficientes del modelo de regresión de rendimiento académico e Uso Internet entretenimiento

		Variables en la ecuación					I.C. 95% para EXP(B)		
		B	E.T.	Wald	gl	Sig.	Exp(B)	Inferior	Superior
Paso 1 ^a	clus_entre2(1)	,349	,239	2,124	1	,145	1,417	,887	2,265
	Constante	-1,340	,138	93,882	1	,000	,262		

a. Variable(s) introducida(s) en el paso 1: clus_entre2.

Fuente: Elaboración propia

El modelo encontrado, se puede expresar de la siguiente forma:

$$\Pr(\text{rendimiento}_{\text{academico}} = 1) = \frac{1}{1 + e^{-(-1,3400,349\text{clusentre2}(1))}}$$

En la Tabla 40 se observa, la que ninguno de los valores de las variable es significativo, los coeficientes de la variable independiente son, Clus_entre2(1)=(OR=1,417, (IC 95% 0,887-2,265),p=0,678)

4.9 Evaluación e Interpretación con el uso de la Herramienta Weka

4.9.1 Resultado Algoritmos para selección de atributos

En la sección de metodología se explica cómo se realizó el procedimiento para la selección de atributos.

A continuación presentamos los resultados del experimento 4 con el que se trabajó, ya que es el que presento menos error cuadrático. Ver Tabla 41.

Tabla 41: Resultados de mejor ajuste de Umbral

EXPERIMENTO 4			
UMBRAL4	1-3		
ERROR	80,7044616	8,9835662	
Cluster	full	Cluster 0	Cluster 1
Instances	297	131	166
Ing	3	3	3
a_for_vir	2	2	2
a_post_aca	3	3	2
a_bus_inf	19	21	18
uso_internet_academico_3	1	2	1

Fuente: Elaboración propia

4.9.2 Resultado del algoritmo SimpleKMeans.

- **Hipótesis Uno**

La vista general del DataSet para esta hipótesis con la clase uso_internet_academico_4 se la puede apreciar en la Figura 27.

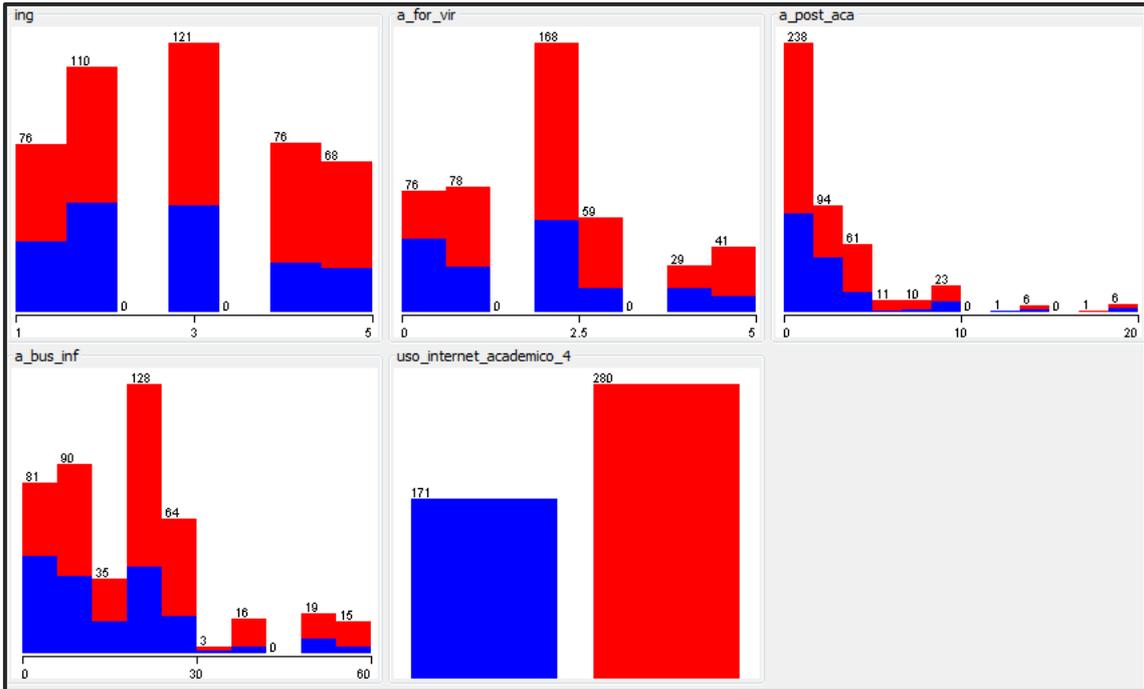


Figura 27: Vista general de las variables que intervienen en la Hipótesis uno

Fuente: Elaboración Propia

Se procede a realizar la comprobación de la primera hipótesis con la herramienta de aprendizaje automático, teniendo en cuenta que se trabaja con la configuración del algoritmo k-means correspondiente al de una semilla de 20 a 25. Con dos cluster, y un máximo de iteraciones de 500. Con el cual se ha obtenido el mínimo número de error cuadrático de los experimentos realizados.

Tras la ejecución del algoritmo *SimpleKMeans* los resultados son los siguientes:

Tamaño de la población: 451 estudiantes

Suma de errores cuadráticos: 121,677974

Instancias de los Cluster:

Cluster 0 → 171 (38%)

Cluster 1 → 280 (62%)

```

kMeans
=====

Number of iterations: 3
Within cluster sum of squared errors: 121.67797371321817
Missing values globally replaced with mean/mode

Cluster centroids:

Attribute                Full Data      Cluster#
                        (451)          0           1
=====
ing                      2.8891         2.7018      3.0036
a_for_vir                2.0222         1.7368      2.1964
a_post_aca               2.6142         2.4971      2.6857
a_bus_inf                19.2905        16.117      21.2286
uso_internet_academico_4 2                1           2

Time taken to build model (full training data) : 0 seconds

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

Clustered Instances

0      171 ( 38%)
1      280 ( 62%)

```

Figura 28: Resultados de la ejecución del algoritmo- SimpleKMeans-Hipótesis uno
Fuente: Elaboración: Propia

En la Figura 28, se puede observar los resultados generados, y sus características o patrones de los estudiantes son:

Grupo 0: Está formado por el 38% de la población estudiantil, como se puede observar su participación en foros virtuales es poca, el tiempo que emplean al mes para realizar tweets o post sobre temas académicos es similar al de grupo 1, el número de horas que emplean los estudiantes para buscar información académica en internet es menor que al otro grupo. Los ingresos económicos de la familia de los estudiantes son bajos en comparación al otro.

Grupo 1: La mayor parte de la población estudiantil se encuentra en este grupo, sus características principales son que tienen mayor participación en foros virtuales, ocupan un tiempo significativo en Twitrear o postear temas referente a su educación, y la actividad en que el estudiante emplea la mayor parte de su tiempo y esfuerzo es en la búsqueda de información académica en Internet.

En los resultados se puede observar que existe una relación directa entre ingreso económico del núcleo familiar y el uso de internet para actividades académicas (Mayor ingreso mayor uso de internet para actividades académicas).

Los estudiantes de acuerdo al nivel de uso de Internet quedarían de la siguiente forma:

Grupo 0: Estudiantes con menor uso de Internet académico.

Grupo 1: Estudiantes con mayor uso de Internet académico.

Si a este análisis de resultados agregamos la variable: nivel uso de Internet académico, podemos observar en la Figura 29, que coincidentalmente corresponde al grupo de usuarios que respondieron a esta pregunta con un valor mayor o igual a 4. Con lo cual se puede deducir que el nivel de ingresos económicos incide directamente en el manejo de internet.

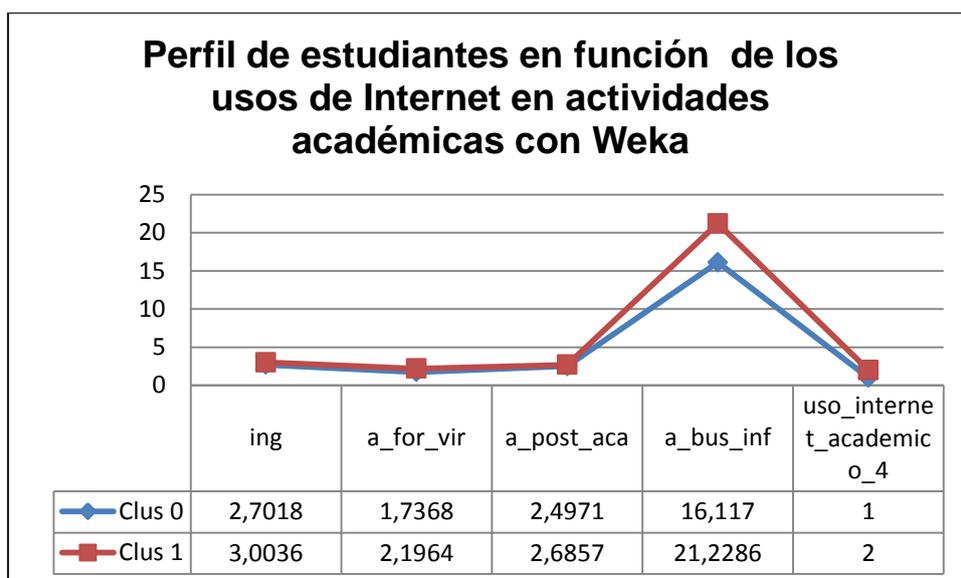


Figura 29: Perfil de estudiantes en función de los usos de Internet en actividades académicas
Fuente: Elaboración propia

Por lo anteriormente expuesto se puede concluir que la hipótesis uno sobre la relación entre los ingresos económicos y el uso de internet para el aprendizaje se cumple, contrastando con los resultados obtenidos en el análisis estadístico realizado en la sección anterior.

- **HIPÓTESIS DOS**

Para estos resultados se realizaron dos análisis, en el primero se utilizaron aquellas variables que fueron seleccionadas utilizando algoritmos de selección de atributos. La vista general nos muestra la Figura 30.

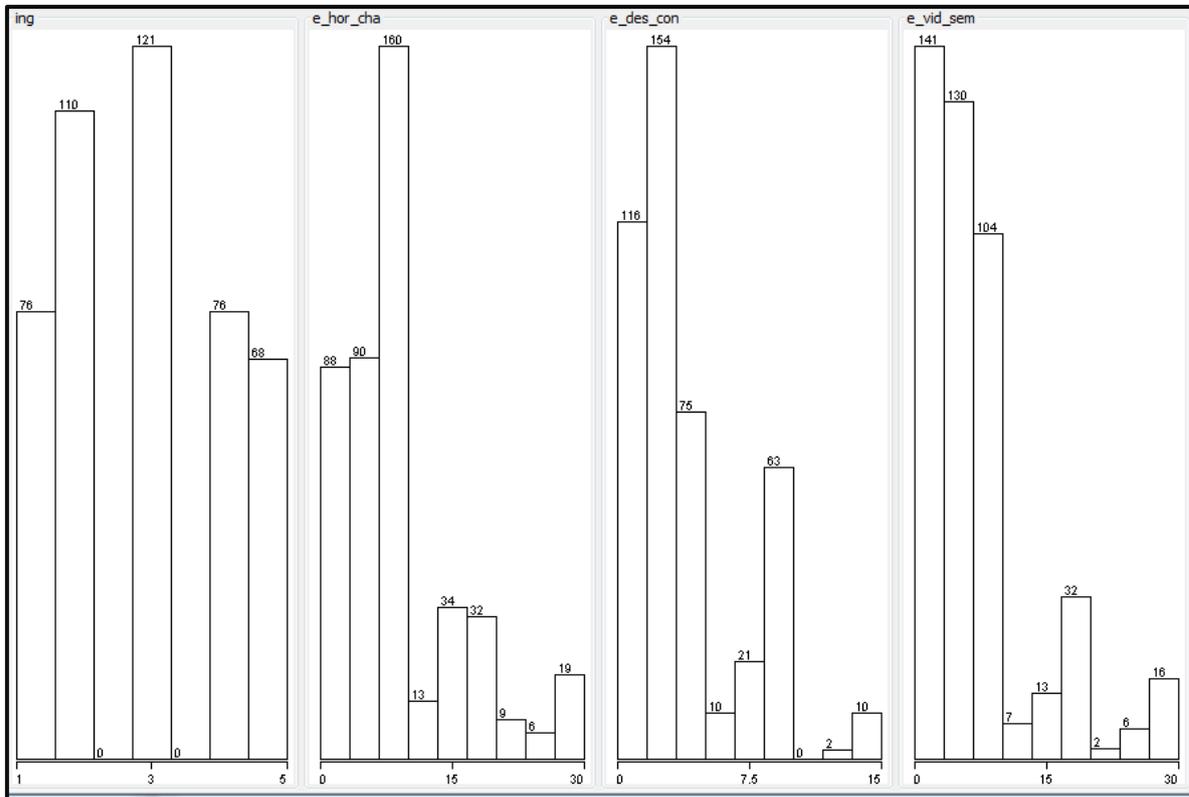


Figura 30: Vista general de las variables que intervienen en la Hipótesis dos-Experimento1

Fuente: Elaboración Propia

Los resultados después de la aplicación del algoritmo *SimpleKMeans* son: ver Figura 31

Tamaño de la población: 451 estudiantes

Suma de errores cuadráticos: 87,1898

Instancias de los Cluster:

Cluster 0 → 239 (53%)

Cluster 1 → 212 (47%)

```

kMeans
=====

Number of iterations: 8
Within cluster sum of squared errors: 87.18978369538519
Missing values globally replaced with mean/mode

Cluster centroids:

Attribute      Full Data      Cluster#
                (451)          0          1
                (239)          (212)
-----
ing            2.8891        1.9331        3.967
e_hor_cha     9.4856        7.4603       11.7689
e_des_con     4.0244        3.0837        5.0849
e_vid_sem     7.6297        5.5607        9.9623

Time taken to build model (full training data) : 0 seconds

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

Clustered Instances

0      239 ( 53%)
1      212 ( 47%)

```

Figura 31: Resultados de la ejecución del algoritmo- SimpleKMeans-Hipótesis dos-experimento 1
Fuente: Elaboración: Propia

En este análisis se pudo observar que efectivamente existe una relación directa entre el ingreso familiar y el uso de internet para entretenimiento.

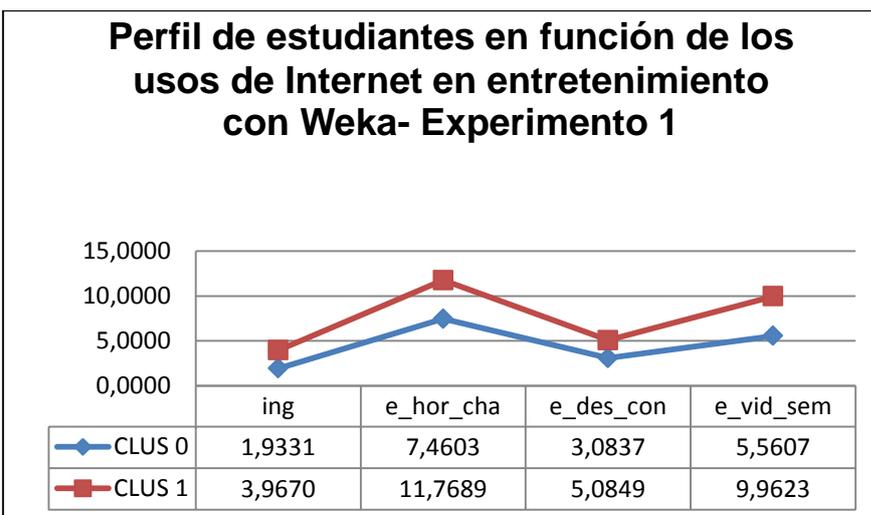


Figura 32: Perfil de estudiantes en función de los usos de Internet en entretenimiento con Weka- Experimento 1
Fuente: Elaboración: Propia

Como se puede observar en Figura 32, las actividades de chatear y ver videos de entretenimiento son los que mayor uso realizan los estudiantes.

En cambio los juegos en línea tienen una menor predilección por parte de los estudiantes, sin embargo, en el grupo de ingresos familiares más alto se puede observar el uso más frecuente de esta actividad.

Sin embargo, las variables obtenidas en la sección anterior de selección de atributos resultaron no ser relevantes para encontrar patrones útiles con el algoritmo de agrupamiento. Esto se puede apreciar fácilmente en la gráfica donde los resultados obtenidos en cada grupo son perpendiculares entre sí, con lo cual no hay forma de obtener conocimiento útil de esta evaluación.

Como el proceso KDD, es iterativo e incremental repetimos la evaluación con el algoritmo SimpleKMeans, pero esta vez considerando todas las variables de entretenimiento, en la cual se obtuvo información más representativa y se pudo sacar patrones, cuyos resultados se presentan a continuación:

La Vista general queda tal como lo indica la Figura 33.

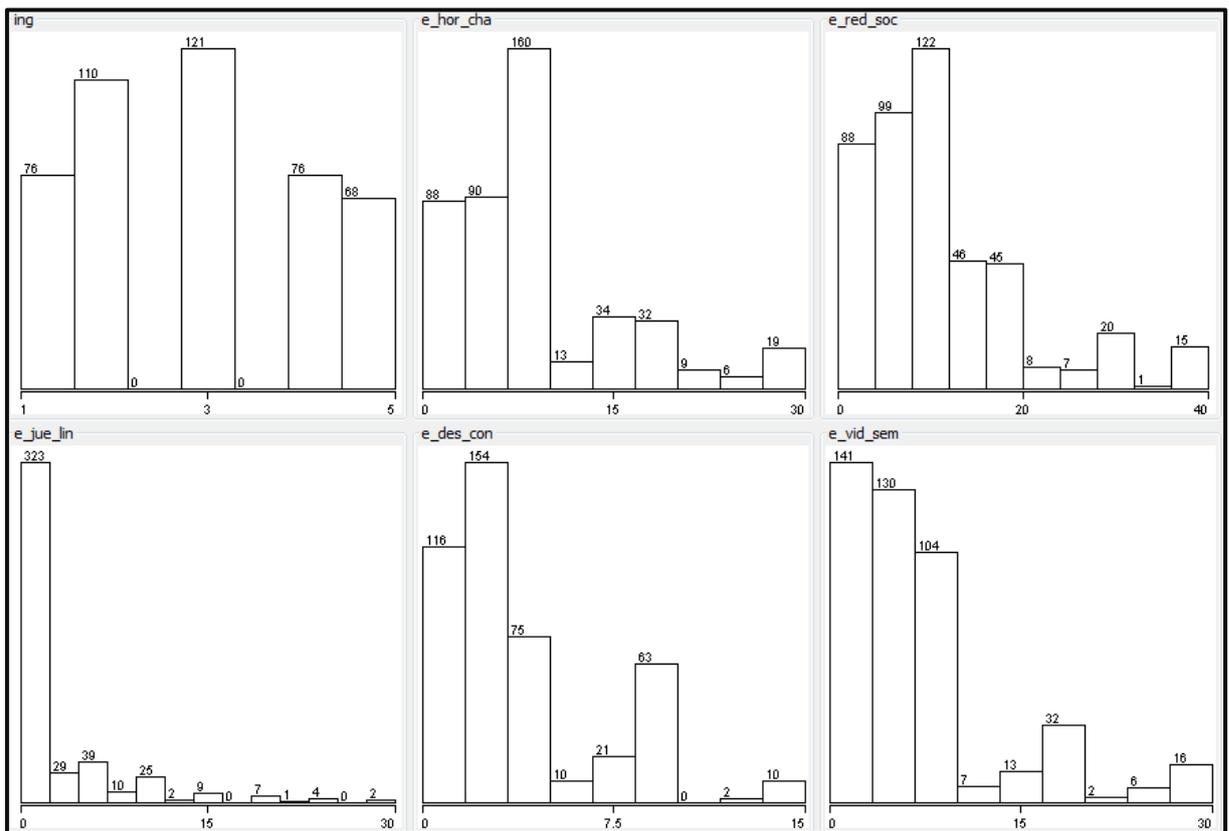


Figura 33: Vista general de las variables que intervienen en la Hipótesis dos-Experimento 2
Fuente: Elaboración Propia

Tamaño de la población: 451 estudiantes

Suma de errores cuadráticos: 119.78936844387368

Instancias de los Cluster:

Cluster 0 → 241 (53%)

Cluster 1 → 210 (47%)

```
kMeans
=====
Number of iterations: 8
Within cluster sum of squared errors: 119.78936844387368
Missing values globally replaced with mean/mode

Cluster centroids:
Attribute      Full Data      Cluster#
                (451)         0           1
                (241)         (210)
=====
ing            2.8891         2.1452      3.7429
e_hor_cha     9.4856         6.1245      13.3429
e_red_soc     11.6718        8.39        15.4381
e_jue_lin     2.6829         1.5768      3.9524
e_des_con     4.0244         2.639       5.6143
e_vid_sem     7.6297         4.7469      10.9381

Time taken to build model (full training data) : 0.01 seconds

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

Clustered Instances

0      241 ( 53%)
1      210 ( 47%)
```

Figura 34: Resultados de la ejecución del algoritmo- SimpleKMeans-Hipótesis dos-experimento 1
Fuente: Elaboración: Propia

A las variables más significativas del análisis anterior (chatear y videos de entretenimiento) se suma el uso de redes sociales y descarga de música y videos. Es así, que se obtuvo las siguientes características:

Grupo 0: Está formado por el 53% de la población estudiantil, menor chat por diversión, Poca utilización de redes sociales, menor número de acceso en línea, realiza poca descarga música, videos y programas, bajas visualizaciones de videos en YouTube cada semana, los Ingresos económicos mensuales de la familia del estudiante son menores que el grupo 1.

Grupo 1: Los estudiantes que se encuentran en este grupo muestran una mayor participación en las redes sociales, chatean activamente, juegan más tiempo en línea, además descargan música, videos y programas a mayor cantidad, los estudiantes pasan gran tiempo observando videos para entretenerse en la cuenta YouTube. Los ingresos económicos de las familias son superiores.

Los perfiles de los estudiantes quedarían representados tal como lo muestra la Figura 35.

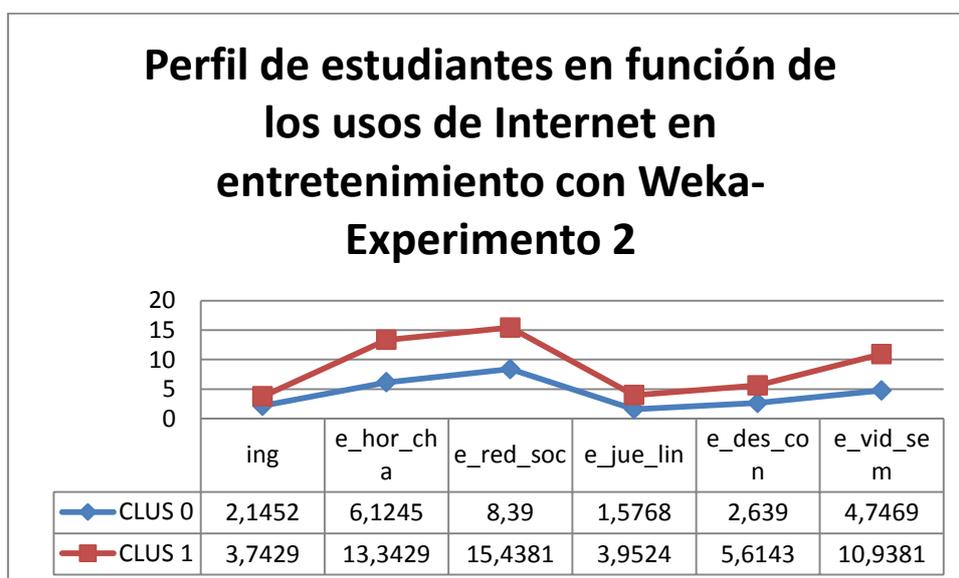


Figura 35: Perfil de estudiantes en función de los usos de Internet en entretenimiento con Weka- Experimento 2
Fuente: Elaboración: Propia

La variable ingresos económicos de la familia de los estudiantes, no es un factor que influye directamente en el uso de Internet para entretenimiento y diversión como se pudo ver en los resultados, es decir, no es un factor para que el estudiantes decida usar el internet o no para este ámbito, ya que la mayor parte de estudiantil utiliza estos recursos.

- **HIPÓTESIS TRES**

A continuación, se detalla el experimento cuatro, ya que es el que posee menos error cuadrático: 51,5824, este experimento utiliza un Seed o raíz de 25.

La vista general del DataSet para esta hipótesis con la clase C_rendimientoAcad se la puede apreciar en la Figura 36.

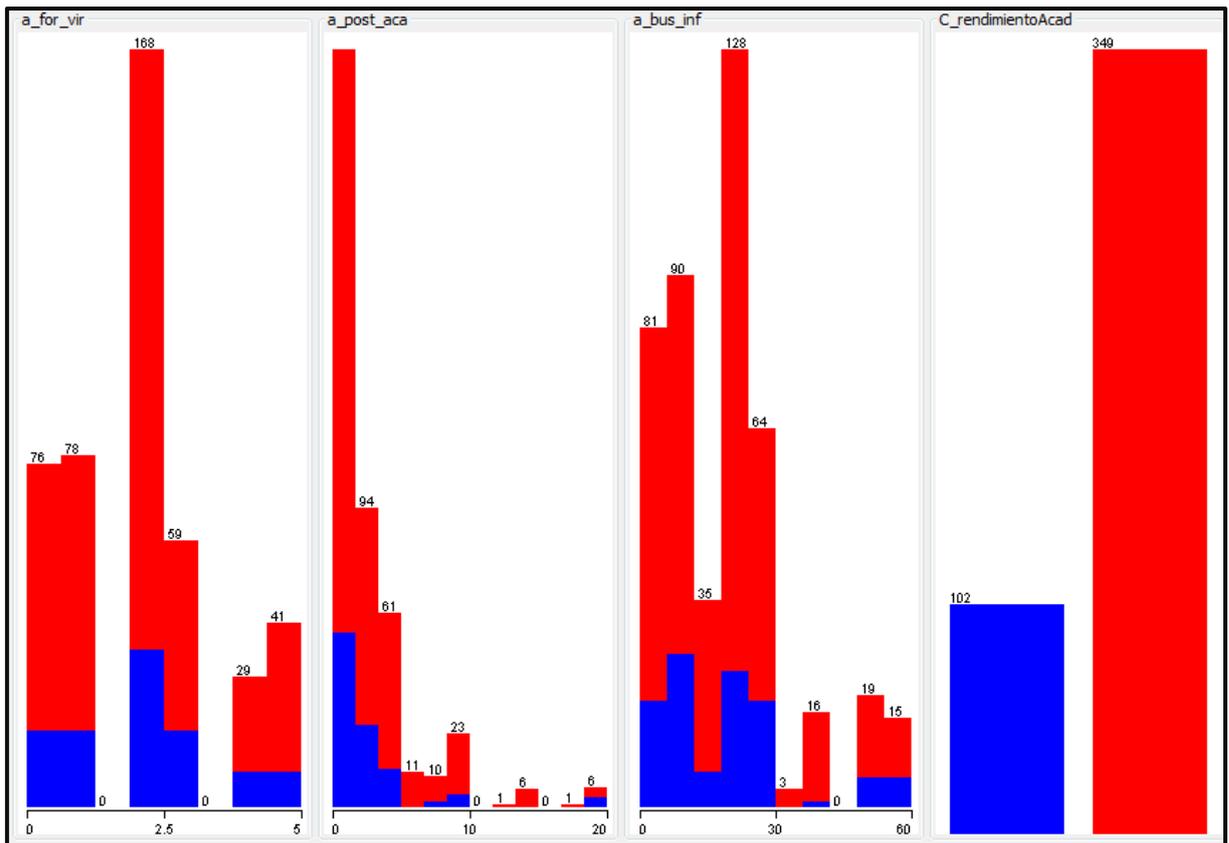


Figura 36: Vista general de las variables que intervienen en la Hipótesis tres

Fuente: Elaboración Propia

Tras la ejecución del algoritmo *SimpleKMeans* los resultados son los siguientes:

Tamaño de la población: 451 estudiantes

Suma de errores cuadráticos: 51.582423065257046

Instancias de los Cluster:

Cluster 0 → 279 (62%)

Cluster 1 → 172 (38%)

Incorrectly clustered instances: 184.0 40.7982 %

```

kMeans
=====

Number of iterations: 13
Within cluster sum of squared errors: 51.582423065257046
Missing values globally replaced with mean/mode

Cluster centroids:

Attribute      Full Data      Cluster#
                (451)         0           1
                (279)         (172)
=====
a_for_vir      2.0222         1.1828      3.3837
a_post_aca     2.6142         1.6201      4.2267
a_bus_inf      19.2905        15.6631     25.1744

Time taken to build model (full training data) : 0.01 seconds

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

Clustered Instances

0      279 ( 62%)
1      172 ( 38%)

Class attribute: C_rendimientoAcad
Classes to Clusters:

  0   1  <-- assigned to cluster
  57  45 | 10
 222 127 | 20

Cluster 0 <-- 20
Cluster 1 <-- 10

Incorrectly clustered instances :      184.0      40.7982 %

```

Figura 37: Resultados de la ejecución del algoritmo- SimpleKMeans-Hipótesis tres
Fuente: Elaboración Propia

En el detalle del experimento en la Figura 37 podemos observar que los integrantes del cluster 1 a pesar de hacer mayor uso de las actividades académicas: post académicos, búsqueda de información y foros virtuales, terminan reprobando al menos una materia.

En cambio los integrantes del cluster 0 que tienen un rendimiento óptimo con 0 materias reprobadas tienen una interacción baja en las actividades académicas consideradas en este estudio, comparados con el cluster 1.

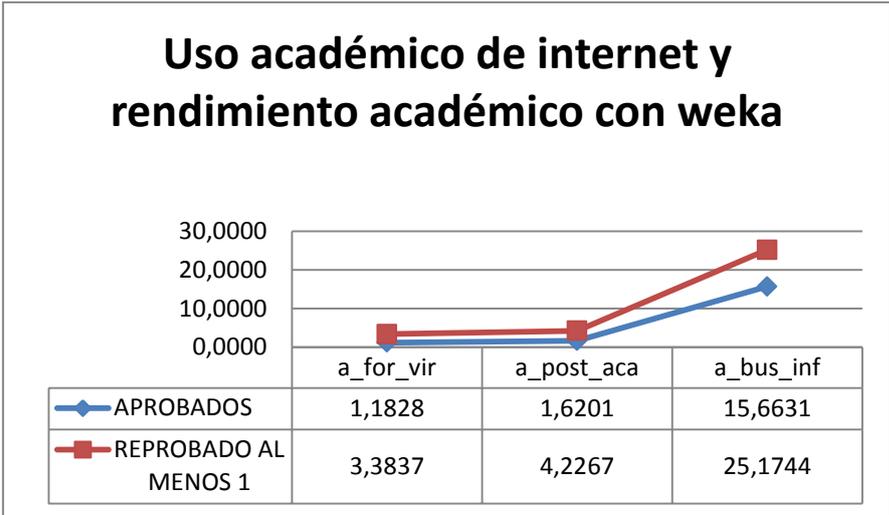


Figura 38: Uso académico de internet y rendimiento académico con Weka
Fuente: Elaboración: Propia

Con estos resultados no se tiene evidencia de alguna relación entre el uso de internet para actividades académicas con el rendimiento final. Ver Figura 38.

- **HIPÓTESIS CUATRO**

Para realizar esta hipótesis se procedió hacer dos análisis: en la primera se consideró aquellas variables obtenidas de la aplicación de filtros de la sección 3.3.3.1.3.

El DataSet en vista general para esta hipótesis con la clase rendimiento_acad se la puede apreciar en la Figura 39.

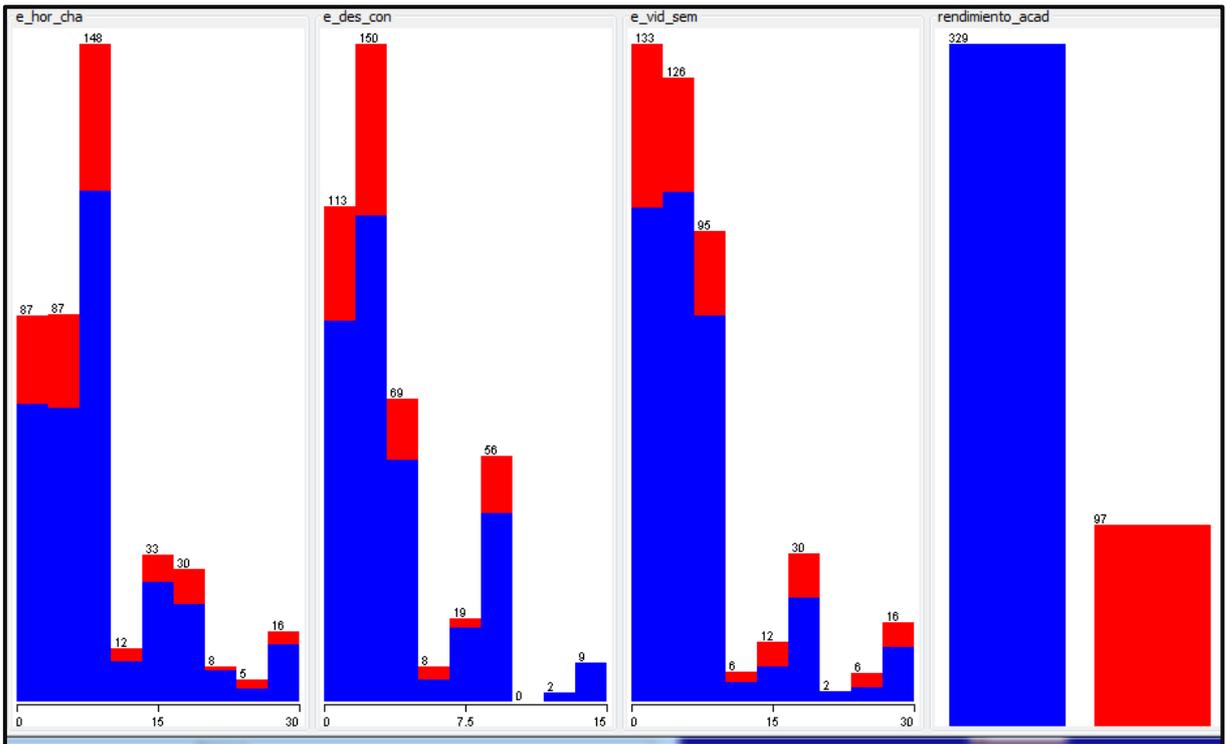


Figura 39: Vista general de las variables que intervienen en la Hipótesis cuatro- Experimento1
Fuente: Elaboración Propia

Los resultados luego del algoritmo SimpleKMeans fueron: Ver Figura 43.

Tamaño de la población: 451 estudiantes

Suma de errores cuadráticos: 142.30627913843603

Instancias de los Cluster:

Cluster 0 → 153 (36%)

Cluster 1 → 273 (64%)

```
kMeans
=====
```

```
Number of iterations: 10
Within cluster sum of squared errors: 142.30627913843603
Missing values globally replaced with mean/mode
```

```
Cluster centroids:
```

Attribute	Full Data (426)	Cluster#	
		0 (153)	1 (273)
e_hor_cha	9.2535	13.4248	6.9158
e_des_con	3.9061	7.0131	2.1648
e_vid_sem	7.6315	13.9216	4.1062
rendimiento_acad	1	1	1

```
Time taken to build model (full training data) : 0.01 seconds
```

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

```
Clustered Instances
```

```
0      153 ( 36%)
1      273 ( 64%)
```

Figura 40: Resultados de la ejecución del algoritmo- SimpleKMeans-Hipótesis cuatro-Experimento1

Fuente: Elaboración Propia

Sin embargo en los experimentos realizados no se presentan evidentes características entre los dos grupos formados por el algoritmo, ya que información relacionada a las variables de análisis no tienen mayores diferencias entre los dos grupos. Esto se puede comprobar fácilmente en el paralelismo que presentan los datos para los dos clusters en la Figura 41.

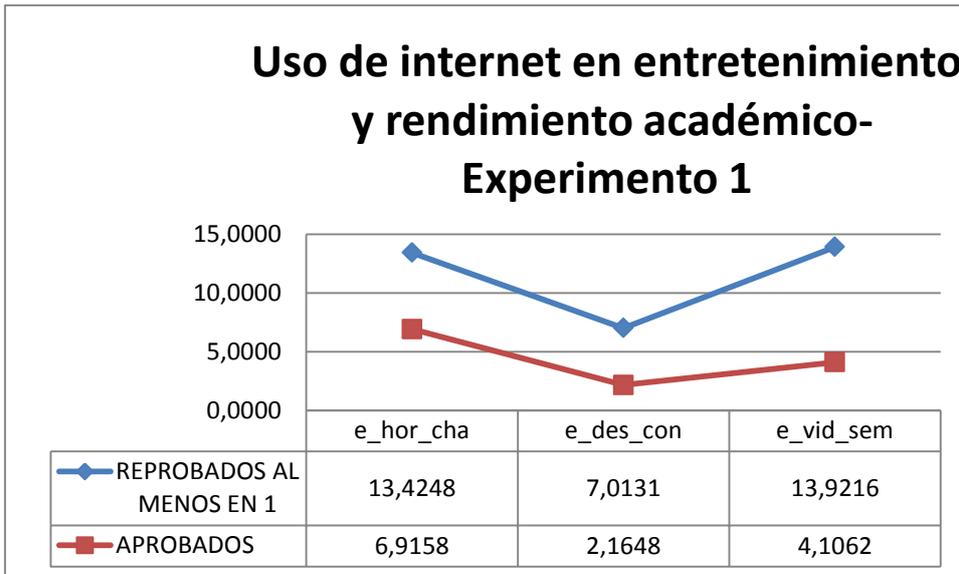


Figura 41: Uso de internet en entretenimiento y rendimiento académico-Experimento 2

Fuente: Elaboración: Propia

En virtud de esto se procedió a realizar un segundo análisis en donde se involucraban todas las variables que tienen que ver con actividades de entretenimiento de los estudiantes.

La vista general del DataSet para esta hipótesis con la clase rendimiento_acad se la puede apreciar en la Figura 42.

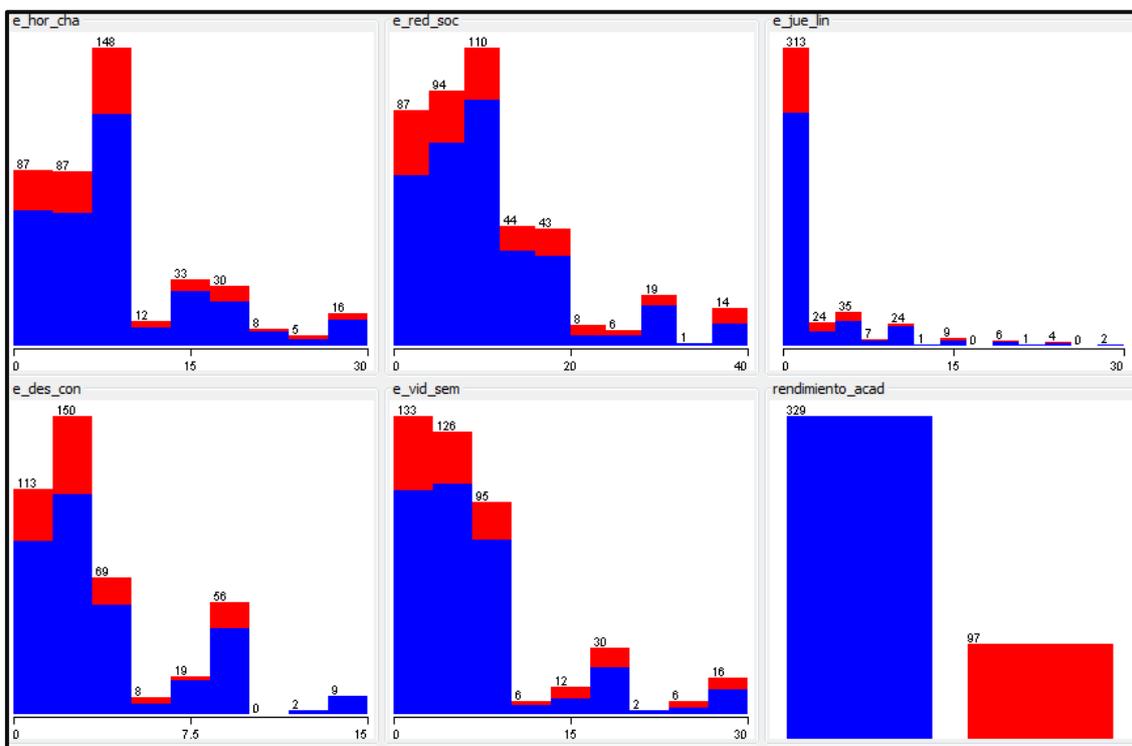


Figura 42: Vista general de las variables que intervienen en la Hipótesis cuatro

Fuente: Elaboración Propia

Tras la aplicación del algoritmo SimpleKMeans los resultados fueron los siguientes: Ver Figura 43.

Tamaño de la población: 451 estudiantes

Suma de errores cuadráticos: 72.6926856241072

Incorrectly clustered instances: 174.0 40.8451 %

Instancias de los Cluster:

Cluster 0 → 273 (64%)

Cluster 1 → 153(36%)

kMeans

=====

Number of iterations: 8

Within cluster sum of squared errors: 72.6926856241072

Missing values globally replaced with mean/mode

Cluster centroids:

Attribute	Full Data (426)	Cluster#	
		0 (273)	1 (153)
e_hor_cha	9.2535	5.967	15.1176
e_red_soc	11.608	7.5824	18.7908
e_jue_lin	2.5869	1.3773	4.7451
e_des_con	3.9061	2.5641	6.3007
e_vid_sem	7.6315	4.8168	12.6536

Time taken to build model (full training data) : 0.01 seconds

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

Clustered Instances

0 273 (64%)

1 153 (36%)

Figura 43: Resultados de la ejecución del algoritmo- SimpleKMeans-Hipótesis cuatro-Experimento 2
Fuente: Elaboración Propia

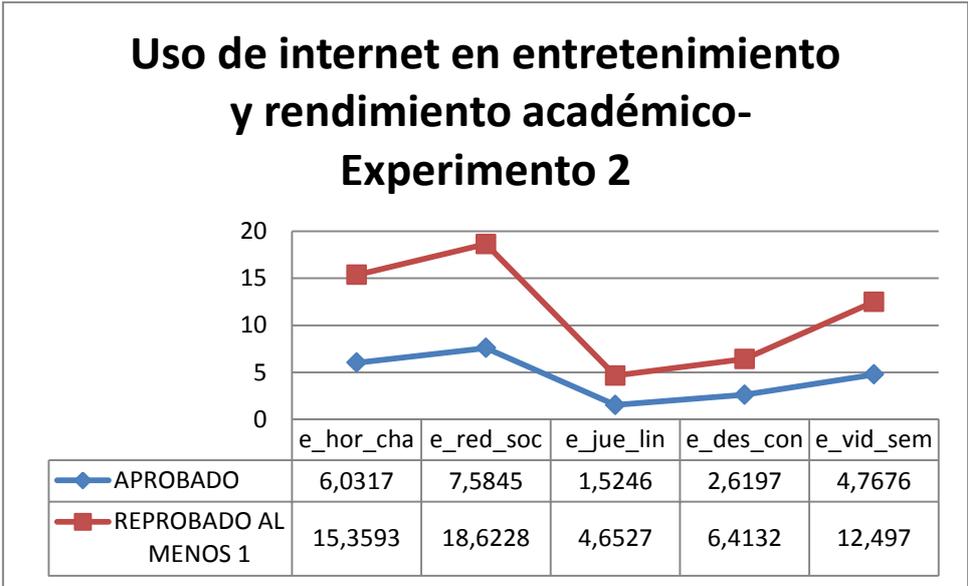


Figura 44: Uso de internet en entretenimiento y rendimiento académico-Experimento 2
Fuente: Elaboración: Propia

Este análisis arrojó datos más significativos; el grupo de estudiantes que dedicó más tiempo en actividades de entretenimiento, tuvieron resultados negativos en su rendimiento académico, en contra parte las personas que dedicaron poco tiempo a estas actividades obtuvieron mejores resultados en su desempeño académico. Ver Figura 44.

En las dos agrupaciones hechas por el algoritmo se observa que el uso de las redes sociales y la visualización de videos de entretenimiento, son las actividades que más se dedican los estudiantes, independiente de su rendimiento académico siendo los videos de entretenimiento los más utilizados

De cara a la comprobación de la hipótesis, se concluyó que efectivamente existe una relación inversa entre el número de horas empleadas en entretenimiento en internet (menos horas de entretenimiento mejores resultados académicos). Con lo cual se comprobaría la hipótesis en cuestión.

CAPÍTULO V: DISCUSIÓN DE RESULTADOS

En esta sección se realiza la discusión de los resultados obtenidos en las diferentes interrogantes planteadas, para lograr los objetivos propuestos.

A fin de establecer conclusiones, sobre la hipótesis “El nivel de ingresos económicos, determina como se utiliza internet para el aprendizaje”, es conveniente aclarar lo siguiente:

La información recopilada de los estudiantes de la Universidad Técnica Particular de Loja, corresponde a una muestra que estuvo estructurada por un 55% de mujeres y un 45% de hombres. De la oferta académica que mantiene la Universidad, la mayor proporción de estudiantes (26,8%) encuestados, corresponde a la titulación de contabilidad y auditoría; mientras que la menor proporción (0,2%), corresponde a la titulación de ingeniería química. Las edades de los sujetos de investigación, estuvieron comprendidas entre 18 y 32 años ($M=21,23$; $DT=2,17$), con una gran concentración de datos por arriba de los 21 años. Por otra parte, en función de la información recopilada se puede afirmar que, los estudiantes provienen de familias con ingresos que se ubican entre 1000 a 1500 dólares mensuales; existiendo una población del 15.1% que perciben más de 1500 dólares por cada mes.

Por otro lado, los resultados obtenidos demuestran que el mayor porcentaje de estudiantes se conecta desde su casa (86%), es decir que existe poco uso del internet en las instalaciones universitarias u otros lugares, y que la frecuencia de conexión son los siete días a la semana, usando en promedio de 4 a 6 horas diarias, con una experiencia de por lo menos cuatro años y un alto grado de conocimiento sobre el uso de las herramientas del internet.

Además de lo expresado, el trabajo investigativo permitió conocer que el promedio de ingreso a la plataforma virtual de la Universidad es de seis horas a la semana, con una desviación estándar de 2,89 horas; usando gran parte del tiempo para realizar intercambio de información o consultas entre compañeros. Como también se evidencia, que realizan alrededor de 5 consultas al profesor de manera directa y no a través de la herramienta virtual.

Así mismo, es necesario resaltar los siguientes resultados:

- Un 39% de los sujetos observan de 3 a 6 videos tutoriales mensuales.
- Poco uso de los tweets académicos, lo cual representa un término medio de tres a mes.
- El promedio del uso del chat en actividades académicas es de 7.59 veces al mes.
- Los estudiantes realizan 19 indagaciones académicas de manera mensual,
- Poco uso de la biblioteca virtual, con promedio de 5,36 horas al mes, con una tendencia a variar superior o inferior a 7,16 en usos no académicos.

Los resultados expresados en el capítulo anterior permitieron conocer también lo siguiente:

- El promedio de chat por diversión es de 9 horas a la semana, con una tendencia a variar a 7 horas.
- Se puede observar que del 23,9%, tan solo el 1.5% de estudiantes no utilizan las redes sociales para el entretenimiento y diversión.
- Se evidencia que el 54,3 % no utiliza los juegos en línea, existiendo un 11% que lo utilizan más de 20 horas y el resto por lo menos 5 horas a la semana.
- Los estudiantes dedican de 4 a 7 horas por semana, para descargar videos de entretenimiento.
- Por otra parte, el trabajo investigativo demuestra el uso de las redes sociales (twitter, facebook y linkelink), por el 100% de la población estudiantil. Y que en su mayoría tienen una gran cantidad de contactos con los cuales interactúan en diferentes actividades.

Al contrastar los resultados observados y que se ha resumido anteriormente, se pudo interpretar que los estudiantes de la Universidad Técnica Particular de Loja, dedican parte de su tiempo en el internet para actividades académicas o aprendizaje, donde la interacción entre estudiantes es el factor más predominante; pero esta herramienta virtual es usada en un mayor grado y de manera general por toda la población estudiantil en actividades de entretenimiento.

Por todo lo expuesto, en referencia a las hipótesis planteadas se puede deducir que, el nivel de ingresos económicos incide en la forma como se utiliza internet para el aprendizaje esto coincide con los estudios de (Zillien & Hargittai, 2009) que hacen referencia a la relación de los Ingresos con el conocimiento, pero no incide en la forma que se utiliza internet para las actividades de entretenimiento. Tal como se demuestra en el apartado 4.8.1.

Con lo referente al rendimiento académico, el proceso investigativo determinó los siguientes resultados:

- Una asociación significativa con la variable de edad (Tau C=0,147, $p=0,001$; $p<0.05$).
- Una incidencia significativa entre el género y el rendimiento académico (Tau B=0,046, $p=0,000$; $p<0.05$)
- El rendimiento académico y el nivel de ingreso no presentan una asociación significativa (Tau C=-0,028, $p=0,511$; $p<0.05$).

Es decir, las actividades direccionadas a los procesos de aprendizaje se relacionan con la edad de las personas. Como también la conducta de cada género, tiene incidencia significativa sobre los resultados del aprendizaje; esto concuerda (Castaño Muñoz, 2010),

que señala que el uso de la tecnología puede causar efectos en el rendimiento académico, a su vez, se concluyó que no existe ninguna inter-relación con el nivel de ingresos.

Al relacionar la edad de las personas con el tiempo de conexión en Internet y su uso, se encontraron los siguientes resultados:

- Se encontró una relación significativa entre edad y el tiempo usado por los estudiantes en el internet (R de Pearson =0,196, $p < 0,00$; $p < 0,05$).
- Así mismo, se determina que existe una relación entre la edad y las horas que el estudiante chatea sobre temas académicos (R de Pearson =-0,116, $p < 0,01$; $p < 0,05$) con un coeficiente de determinación del 1,34%.
- Hay una relación significativa entre la edad y las asignaturas en las cuales el estudiante se matriculó en el semestre anterior (R de Pearson =-0,254, $p < 0,00$; $p < 0,05$) y las que reprobó (R de Pearson =-0,299, $p < 0,000$; $p < 0,05$). La dependencia de la variable edad sobre la variable matrícula alcanza un coeficiente de determinación de 6,45 %, y de la variable que aprobó un 8,94 %.

Si se analiza esta información con los resultados anteriores, nacen otras hipótesis que pueden ser consideradas para investigaciones futuras como son:

- Existe una significativa relación entre la edad de las personas y el uso del internet, de tal manera que, se puede afirmar que personas con una mayor edad usan esta herramienta virtual para investigar y aplicar en sus procesos de aprendizaje.
- El género femenino utiliza con mayor frecuencia el internet como una herramienta virtual, para sus actividades académicas.

Con respecto a los ingresos y su incidencia en el uso del internet, los resultados obtenidos demuestran que:

- Una asociación significativa entre los ingresos y el lugar de conexión ($\chi^2 = 30,21$, $p < 0,017$; $p < 0,05$).
- Existe una relación entre la variable ingresos y el tiempo usando para la conexión (R de Pearson =0,226, $p < 0,000$; $p < 0,05$), con un coeficiente de determinación del 5,10 %.
- Asimismo, se encontró una relación significativa entre ingresos y el nivel de conocimiento en el manejo de internet (R de Pearson =0,228, $p < 0,000$; $p < 0,05$). Con un coeficiente de determinación del 5,1984%.

Es decir que, un mayor nivel de ingresos permite seleccionar el lugar y el tiempo dedicado a la conexión. De por sí una mayor dedicación al uso de esta herramienta incrementa el

nivel de conocimiento sobre el mismo, pero no hay ninguna incidencia del nivel de ingresos con respecto al rendimiento académico.

El siguiente análisis parte de las actividades académicas y luego de haberlos organizados en diferentes clases se obtuvieron en resumen los siguientes resultados:

- El primer grupo denominado modernos, está conformado por el 23,9% del total analizado usan con mayor frecuencia los recursos académicos de la plataforma virtual, y así como también realizan consultas entre compañeros a través del chat académicos. La edad promedio de este grupo es de 21 años.
- El grupo 2, incluye al 76,1% de estudiantes utilizan muy poco los recursos de internet como medio de aprendizaje. La edad promedio de este grupo es de 21,37 años.

Por otra parte, los resultados obtenidos sobre el uso de internet en actividades de entretenimiento, a partir de la conformación de dos grupos y al investigar sobre el uso del Internet en chat, redes sociales y videos por entretenimiento, se destaca lo siguiente:

- El grupo moderno incluye al 29,5% de estudiantes, su principal características es el mayor uso de las redes sociales y el de chat por diversión, chatear y además de observar videos académicos. La edad promedio de este grupo es de 21,31.
- El grupo pasivos constituye el 70,5% de estudiantes, este grupo utiliza muy poco los recursos de internet como medio de entretenimiento y diversión. La edad promedio del grupo pasivo es de 21,20.

En este análisis los estudiantes fueron seleccionados y agrupados, luego del trabajo investigativo realizado se encontró:

El grupo 1, llamado tradicional y conformado por el 40,1% de estudiantes, son los que utilizan en menor grado estos dispositivos electrónicos. La edad promedio es de 21,49 años.

El grupo llamado moderno y que lo conforman el 59,9% de estudiantes, tienen un mejor acceso al uso de estos dispositivos. La edad promedio de este grupo es de 21,06 años.

Un 60% de la población estudiantil tienen acceso a estos tipos de dispositivos electrónicos, lo que les permiten poder utilizar las herramientas del internet en cualquier hora y lugar, desde el cual quiere acceder. Por ende, es que la comunicación o el uso de esta herramienta son más frecuentes en los domicilios de las personas

Con estos datos se pudo realizar las regresiones logísticas con los perfiles de cada estudiante, con el fin de comprobar las hipótesis, que a continuación se explicara.

Análisis de la comprobación de las hipótesis

Perfil Académico.

De la información anteriormente analizada y mediante la prueba de contraste para la hipótesis Uno. Se obtuvo significativo ($\chi^2=15,999$, $p=0,042$; $p<0.05$); se puede afirmar que existe relación entre los niveles de ingresos económicos y el tiempo destinado al uso del internet.

Para la verificación de las hipótesis, sobre la incidencia del internet como una herramienta para el aprendizaje, al utilizar las diferentes pruebas estadísticas se pudo encontrar lo siguiente:

No todas las variables alcanzan un valor significativo ($p<0,05$). La puntuación de Wald para el modelo probado indica que la variable uso_internet_académico aporta significativamente a la predicción de la variable ingreso, los resultados obtenidos se pueden generalizar a la población.

El odd ratio (OR) de pertenecer al perfil de tradicional con respecto al perfil moderno es de 3,181 veces mayor cuando el estudiante pertenece al nivel de ingresos 3, es decir, de hasta 1000 dólares, respecto al nivel 1 (OR=3,181, (IC 95% 1,557-3,860), $p=0,01$); es de 1,931 veces mayor si el estudiante pertenece al nivel de ingreso 4, es decir, de hasta 1.500 dólares, respecto al 1 (OR=1,931, (IC 95% 1,009-3,696), $p= 0,047$).

Con los valores encontrados se puede señalar que los estudiantes con mayores ingresos tienen más posibilidades de utilizar el internet para actividades académicas.

De los experimentos realizados con la herramienta Weka anteriormente se puede concluir, que la hipótesis uno sobre la relación entre los ingresos económicos y el uso de internet para el aprendizaje se cumple, contrastando con los resultados obtenidos en el análisis estadístico.

En función de estos dos resultados se pueden afirmar, que el nivel de ingresos si inciden en la capacidad que tiene un estudiante para acceder al uso del internet y por ende usar esta herramienta en las actividades de aprendizaje, concordando con (Zillien & Hargittai, 2009) que proponen que existe relación entre los ingresos económicos con los conocimientos que adquieren los estudiantes.

Perfil de Entretenimiento

Al aplicar las pruebas de contrastes de la hipótesis 2, que sustenta que el nivel de ingresos determina como se utiliza internet para entretenimiento. Mediante la técnica estadística chi-cuadrado de Pearson, se obtuvo como resultados ($\chi^2=9,522$, $p=0,300$);. Con este valor de significancia ($P < 0,05$), se concluyó que el uso de internet para entretenimiento no depende de los ingresos.

Para el análisis y contraste de esta hipótesis, se procedió a realizar un modelo de regresión logística binaria, los resultados de la prueba ómnibus, indican que el uso de internet para entretenimiento no alcanzan un valor significativo ($X^2=4,401$, $p=0,355$; $P < 0,05$). El estadístico R2 Nagelkerke con valor de 0,014 (1,4%), La prueba de Hosmer y Lemeshow, $X^2=0,000$, $p=1,000$), el valor de $p=1,000$ indica que se acepta H_0 , es decir, que valores observados son iguales a los encontrados, se considera el modelo.

Con el análisis Weka, en cambio se ven resultados significativos comprobando que a mayor ingresos mayor uso de internet en actividades de entretenimiento, aquí infiere un poco con los datos estadísticos obtenidos, sin embargo nos quedamos con los resultados realizados con fórmulas estadísticas, ya que son resultados con más aciertos, esto conduce a concluir que el uso de internet no incide en el rendimiento académico de los estudiantes, es decir, que el modelo no se ajusta a los datos, que el nivel de ingresos económicos no inciden en la capacidad que tiene un estudiante para acceder al uso del internet para el entretenimiento.

Usos académicos de internet y rendimiento académico

Se contrasta adicionalmente la hipótesis 3, que establece que el uso de la tecnología en el aprendizaje incide en el rendimiento académico.

Los indicadores muestran que el rendimiento académico, no incide con el uso de Internet académico, esto explica que la significancia del modelo es mayor a 0,05, es decir, el modelo no es significativo. ($R^2=0,171$, $p < 0,680$; $p < 0,05$). Hosmer y Lemeshow no da un valor de significancia esto implica que el modelo se ajusta a los datos.

Con el análisis Weka, esta hipótesis no se comprueba por que el grupo de estudiantes que tienen un rendimiento óptimo, utilizan menos el internet para fines académicos, que las personas del grupo 2, que terminan reprobando al menos una materia y no demuestran una relación directa entre las dos variables.

Por ende, con estos dos resultados obtenidos, se puede afirmar que el uso de tecnología en el aprendizaje no es un factor influyente en el rendimiento académico de los estudiantes.

De igual manera, al analizar la hipótesis 4 sobre el uso del internet en actividades de entretenimiento y su incidencia en el rendimiento académico. Con $X^2= 2,085$, se obtiene una probabilidad de $p=0,149$, el estadístico R2 Nagelkerke, el cual nos da un valor de 0,07 que equivale al 7% de la varianza.

En el análisis Weka, se observa que hay una relación inversa entre el rendimiento académico y las actividades de entretenimiento, las personas que utilizan su tiempo en actividades de entretenimiento no tienen un rendimiento académico óptimo.

Con estos resultados obtenidos a base de dos herramientas se establece que el uso de la tecnología para entretenimiento no incide en el rendimiento académico de los estudiantes de la Universidad Técnica Particular de Loja.

CONCLUSIONES

A más de las conclusiones expuestas a lo largo de este trabajo, y con los resultados obtenidos con las técnicas de minería de datos, se indican las siguientes:

- El 100 % de los estudiantes encuestados de tercer a décimo ciclo de la Universidad Técnica Particular de Loja, emplean el Internet para el desarrollo de sus actividades académicas y de entretenimiento.
- Las técnicas estadísticas y minería de datos tienen importantes ventajas, para descubrir patrones de comportamiento en el uso de Internet de los estudiantes, ya que asocian una serie de elementos que permiten ejecutar una mejor validación experimental del modelo y un análisis de resultados más completo y fiable.
- El clustering al categorizarse como descriptivo, fue la técnica que mejor se adaptó, para la realización de esta investigación; brindando resultados eficaces referentes a la incidencia de la variable ingresos económicos de las familia de los estudiantes y el uso del Internet, ya que, con esta técnica se pudo conocer las características principales mediante perfiles, y a su vez, permitió relacionar el rendimiento académico con el Uso de Internet.
- La técnica de K means, fue la más adecuada para emplear en esta investigación, por ejecutar la conformación de conglomerados cuando el tamaño de la muestra no es de gran tamaño, como en este caso.
- El nivel de ingresos económicos de la familia del estudiante incide en el uso de Internet académico ($R^2=15,999$, $p=0,042$; $p<0.05$), como se observa esta incidencia es baja; sin embargo, está se refleja más en las familias de los estudiantes que oscilan entre 350 y 600 dólares, mientras que en los niveles superiores va desapareciendo (3,4,5), comprobando así la hipótesis uno, que sostiene que el nivel de ingresos económicos determina como se utiliza el internet para el aprendizaje, así mismo; después de analizar los resultados de Weka, se concluye que la hipótesis se cumple, contrastando con los resultados obtenidos en el análisis estadístico realizado en la sección anterior.
- Con los experimentos realizados para comprobar la hipótesis tres, que manifiesta que el uso de internet en el aprendizaje incide en el rendimiento académico del

estudiante. Se evidencia que no existe una relación significativa ($R^2=0,171$, $p<0,680$; $p<0.05$), lo cual se demostró mediante regresión logística y la herramienta Weka; pero es importante destacar que los estudiantes poseen la experiencia y la capacidad necesaria, para realizar las actividades académicas, pero no hacen el uso adecuado de las mismas.

- Por lo antes expuesto en la literatura, se pudo observar que existen indicios de que el uso del Internet en actividades de entretenimiento y diversión inciden negativamente en el rendimiento académico del estudiante; sin embargo, en esta investigación no se pudo encontrar evidencias para afirmar este argumento, es decir, el uso de Internet en entretenimiento, no incide significativamente sobre el rendimiento académico ($R^2=2,085$, $p<0,149$; $p<0.05$) de los estudiantes de la UTP.
- El uso de Internet, tanto académico como de entretenimiento, no incide en el rendimiento académico, esto indica que se debe replantear la utilización de las tecnologías a manera que el estudiante pueda hacer uso de las mismas como parte de sus actividades de aprendizaje, siempre buscando un resultado positivo.
- Al utilizar dos herramientas para realizar la minería de datos, se comprueba la autenticidad de la información, esto significa que se obtienen los mismos resultados enfocados a las comprobaciones de las hipótesis.

REFERENCIAS

- Alcalay, L., & Antonijevic, N. (1987). *Motivación para el aprendizaje : variables afectivas. Revista de educación (Santiago)*. Chile.
- Álvarez Sáiz, E., & Zorrilla Pantaleón, M. E. (2012). Learning styles and academic performance in college courses learning tool oriented computer taught in virtual form. *Estilos de aprendizaje: investigaciones y experiencias: [V Congreso Mundial de Estilos de Aprendizaje], Santander, 27, 28 y 29 de junio de 2012*. Retrieved from <http://dialnet.unirioja.es/servlet/articulo?codigo=4636887&info=resumen&idioma=SPA>
- Arnal, J., Rincón, D. del, & Latorre, A. (1992). *Investigación educativa : fundamentos y metodologías*. (R. Ilustrada, Ed.).
- Ayala, C. (2007). Relación entre el Uso de Internet y el Logro Académico.
- Ayala, C. (2007). *Tesis Relación entre el Uso de Internet y el Logro Académico*.
- Blumler, J. G., & Katz, E. (1974). *The Uses of mass communications: current perspectives on gratifications research, Volume 1974, Part 1*. Sage Publications. Retrieved from http://books.google.com.ec/books/about/The_Uses_of_mass_communications.html?id=Op7YAAAAIAAJ&pgis=1
- Castaño Muñoz, J. (2010). *La desigualdad digital entre los alumnos universitarios de los países desarrollados y su relación con el rendimiento académico. Revista de Universidad y Sociedad del Conocimiento*.
- Castaño Muñoz, J. (2011). El Uso De Internet Para La Interacción En El Aprendizaje : Un Análisis De La Eficacia Y La Igualdad.
- Cho, C.-M. (2004). How to measure the digital divide? Retrieved January 10, 2015, from <https://www.itu.int/osg/spu/ni/digitalbridges/presentations/02-Cho-Background.pdf>
- Clark, P., & Robin, B. (2000). *Data Mining. Practical Machine Learning Tools and Techniques with java implementations*. (M. Kaufmann, Ed.).
- Coll, C. (2011). Psicología de la educación y prácticas educativas mediadas por las tecnologías de la información y la comunicación. Retrieved January 7, 2015, from <http://virtualeduca.org/ifdve/pdf/cesar-coll-separata.pdf>
- CONATEL. (2013a). Conatel aprueba regulación para el acceso a internet | Secretaría Nacional de Telecomunicaciones | Ecuador. Retrieved January 14, 2015, from <http://www.regulaciontelecomunicaciones.gob.ec/conatel-aprueba-regulacion-para-el-acceso-a-internet/>
- CONATEL. (2013b). REFORMAS AL REGLAMENTO PARA PRESTAR EL SVA (156). Retrieved January 14, 2015, from <http://www.regulaciontelecomunicaciones.gob.ec/wp-content/uploads/downloads/2013/11/TEL-595-26-CONATEL-2013-REFORMAS-AL-REGLAMENTO-PARA-PRESTAR-EL-SVA2.pdf>
- Cuneo, C. (2002). Globalized and Localized Digital Divides Along the Information Highway: A Fragile Synthesis Across Bridges, Ramps, Cloverleaves, and Ladders. Retrieved

- January 12, 2015, from http://experiential-ed.mcmaster.ca/institute-on-globalization-and-the-human-condition/documents/IGHC-WPS_02-2_Cuneo.pdf
- Dader, J. L. (1992). *“Teorías de transición”*, en VV.AA. *Opinión pública y comunicación política* (Eudema.). Madrid.
- DiMaggio, P., & Hargittai, E. (2001). From the “digital divide” to “digital inequality”: Studying Internet use as penetration increases. ... *University Center for Arts and Cultural* ... Retrieved from http://www.maximise-ict.co.uk/WP15_DiMaggioHargittai.pdf
- DiMaggio, P., Hargittai, E., Celeste, C., & Shafer, S. (2004). From Unequal Access to Differentiated Use: A Literature Review and Agenda for Research on Digital Inequality. Retrieved January 9, 2015, from <http://www.webuse.org/pdf/DiMaggioEtAl-DigitalInequality2004.pdf>
- Duart, J. M. (2010). Nuevas brechas digitales en la educación superior. Retrieved January 13, 2015, from http://journals.uoc.edu/index.php/rusc/article/viewFile/v7n1_editorial/v7n1_editorial_esp
- El Comercio. (, August 17). El uso de Internet en Ecuador creció 11 veces en siete años | El Comercio. 2014. Retrieved from <http://www.elcomercio.com.ec/tendencias/ecuador-internet-datos-tecnologia-usuarios.html>
- Farias, R. A., Durán, E. B., & Figueroa, S. G. (2008). Las Técnicas de Clustering en la Personalización de Sistemas de e-Learning. Retrieved August 10, 2015, from http://sedici.unlp.edu.ar/bitstream/handle/10915/21990/Documento_completo.pdf?sequence=1
- Forteza Méndez, J. A. (1975). *Modelo instrumental de las relaciones entre variables motivacionales y rendimiento*. *Revista de Psicología General y Aplicada*. (Á. Editorial, Ed.). España.
- Godoy, S., & Herrera, S. (2003). ¿Qué ocurre cuando se usa (y no se usa) Internet? Resultados del World Internet Project-Chile. *Cuadernos de Información*. Retrieved from <http://cuadernos.uc.cl/uc/index.php/CDI/article/view/161>
- Hernández, J., Ramírez, M., & Ferri, C. (2007). *Introducción a la Minería de Datos*. (P. Education, Ed.). México.
- INEC. (2011). *REPORTE ANUAL DE ESTADÍSTICAS SOBRE TECNOLOGÍAS DE LA INFORMACIÓN Y COMUNICACIONES (TIC'S) 2011*. Retrieved from http://www.inec.gob.ec/sitio_tics/presentacion.pdf
- INEC. (2012). *Tecnologías de la Información y Comunicaciones (TIC'S) 2012*. Retrieved from http://www.inec.gob.ec/sitio_tics2012/presentacion.pdf
- INEC. (2013a). Empleo (Encuesta Nacional de Empleo, Desempleo y Subempleo-ENEMDU) | Instituto Nacional de Estadística y Censos. Retrieved January 14, 2015, from <http://www.ecuadorencifras.gob.ec/empleo-encuesta-nacional-de-empleo-desempleo-y-subempleo-enemdu/>
- INEC. (2013b). *Tecnologías de la Información y Comunicaciones (TIC'S) 2013*. Retrieved

- January 14, 2015, from http://www.ecuadorencifras.gob.ec/documentos/web-inec/Estadisticas_Sociales/TIC/Resultados_principales_140515.Tic.pdf
- ITU. (2005). BUILDING DIGITAL BRIDGES. Retrieved from http://www.itu.int/osg/spu/ni/wsisbridges/linked_docs/Building_exec_A5_fin-e.pdf
- Jiménez, M. V. G., Izquierdo, J. M. A., & Blanco, A. J. (2000). La predicción del rendimiento académico: regresión lineal versus regresión logística. Retrieved from <http://www.psicothema.com/pdf/558.pdf>
- Katz, E., Blumer, J. G., & Gurevitch, E. M. (1973). *Uses and gratifications research. The Public Opinion Quarterly* (Praeger Pu.). Nueva York.
- LaRose, R., Mastro, D., & Eastin, M. S. (2001). Social Science Computer Review. Retrieved January 13, 2015, from <http://www.ensani.ir/storage/Files/20110209150302-%D9%81%D9%87%D9%85%DA%A9%D8%A7%D8%B1%D8%A8%D8%B1%D8%AF%D9%87%D8%A7%DB%8C%D8%A7%DB%8C%D9%86%D8%AA%D8%B1%D9%86%D8%AA.pdf>
- Maggio, M. (2007). *Diálogos en educación, entrevistas: Buenos Aires*. Buenos Aires.
- National Telecommunications & Information Administration (NTIA). (1995). Falling through the net: a survey of the “have nots” in rural and urban America. Retrieved January 8, 2015, from <http://www.ntia.doc.gov/>
- Pippa Norris. (2001). *Digital divide: civic engagement, information poverty and the Internet worldwide*. (C. University, Ed.). Cambridge.
- Proenza, F. J., Bastidas-Buch, R., & Guillermo, M. (2001). Telecentros para el desarrollo socioeconómico y rural en América Latina y el Caribe. Retrieved from http://www.itu.int/ITU-D/ict/mexico04/doc/doc/10_Telecenters_s.pdf
- Rodríguez Morales, F. J.--frarodri@alumnos. ubiobio. c. (2015, June). Aplicación de técnicas de data mining al rendimiento académico del primer curso de programación de computadores.
- Serrano, A., & Martinez, E. (2003). *La Brecha Digital: Mitos y Realidades*. (UABC, Ed.) Mexico. Mexico.
- Sorice, M. (2005). *I media. La prospettiva sociologica*. (Carocci, Ed.). Roma.
- SUPERTEL, S. de T. (2010). Aspectos que influyen en los costos del servicio de Internet. Retrieved from http://www.supertel.gob.ec/pdf/publicaciones/revista_supertel_7.pdf
- Taylor Wallace, Dekkers, J., Marshall, S., & Zhu, G. X. (2003). Socio-economic Factors Affecting Home Internet Usage Patterns in Central Queensland. Retrieved January 8, 2015, from <http://www.inform.nu/Articles/Vol6/v6p233-246.pdf>
- Teknomos, K. (2007). Ejemplo numérico de Agrupamiento K-medias. Retrieved August 10, 2015, from <http://people.revoledu.com/kardi/tutorial/kMean/EjemploNumerico.htm>
- Torres Díaz, J. C. (2012). *Análisis de las relaciones entre los niveles de ingreso, edad y género de los estudiantes, los usos de internet y el rendimiento académico en un grupo de universidades ecuatorianas presenciales*.

- UIT. (2004). Building Digital Bridges. Retrieved January 10, 2015, from <https://www.itu.int/osg/spu/ni/digitalbridges/>
- UNESCO. (2005). *Hacia las sociedades del conocimiento*. (UNESCO, Ed.). 7 Place de Fontenoy; 75732 Paris 07 SP. Retrieved from <http://unesdoc.unesco.org/images/0014/001419/141908s.pdf>
- UNESCO. (2009). Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura. Retrieved January 12, 2015, from <http://www.unesco.org/new/es/>
- Unión Internacional de Telecomunicaciones. (n.d.). UIT: Comprometida para conectar el mundo. Retrieved January 10, 2015, from <https://www.itu.int/es/Pages/default.aspx>
- van Dijk, J., & Hacker, K. (2003). The Digital Divide as a Complex and Dynamic Phenomenon. Retrieved January 10, 2015, from <http://doc.utwente.nl/58689/1/Dijk03digital.pdf>
- Zillien, N., & Hargittai, E. (2009). Digital Distinction: Status-Specific Types of Internet Usage. *Social Science Quarterly*, 90(2), 274–291. doi:10.1111/j.1540-6237.2009.00617.x

ANEXOS

Anexo 1

Encuesta aplicada a los estudiantes

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

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

2. Responda la siguiente pregunta	
¿Qué carrera estudia?	

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

4. Responda la siguiente pregunta	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 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		máximo
¿Cuántas veces a la semana ingresa a la plataforma virtual de su universidad?	(____)	14
¿Aproximadamente cuántas consultas les hace a sus profesores cada mes?	(____)	30
¿Aproximadamente cuántas consultas realiza a sus compañeros por mes?	(____)	30
¿Aproximadamente cuántos recursos educativos descarga de la plataforma virtual cada mes?	(____)	20
¿Aproximadamente cuántos videos académicos mira en YouTube cada mes?	(____)	20
¿Aproximadamente en cuántos foros virtuales participa cada mes?	(____)	5
¿Aproximadamente cuántos post o tweets sobre temas académicos realiza en las redes sociales por mes?	(____)	20
¿Aproximadamente cuántas horas chatea sobre temas académicos cada mes?	(____)	20
¿Aproximadamente cuántas horas busca información académica en internet cada mes?	(____)	60
¿Aproximadamente cuántas horas utiliza la biblioteca virtual de la universidad cada mes?	(____)	30

11. En lo referente al entretenimiento y diversión en internet	
¿Aproximadamente cuántas horas a la semana chatea por diversión?	(____)30
¿Aproximadamente cuántas horas a la semana utiliza redes sociales?	(____)40
¿Aproximadamente cuántas horas a la semana utiliza juegos en línea?	(____)30
¿Aproximadamente cuántas horas a la semana descarga música, videos y programas?	(____)15
¿Aproximadamente cuántos videos para entretenimiento mira en YouTube cada semana?	(____)30

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

Perfil de estudiantes en función las actividades académicas

Análisis k medias para clasificar en dos grupos a los estudiantes en función las actividades académicas.

Centros de los conglomerados finales

	Conglomerado	
	1	2
¿Cuántas veces a la semana ingresa a la plataforma virtual de su universidad?	6,33	5,46
¿Aproximadamente cuántos recursos educativos descarga de la plataforma virtual cada mes?	8,81	5,98
¿Aproximadamente cuántas horas chatea sobre temas académicos cada mes?	10,07	6,81
¿Aproximadamente cuántos videos académicos mira en YouTube cada mes?	8,73	5,43
¿Aproximadamente cuántas horas busca información académica en internet cada mes?	38,99	13,09
¿Aproximadamente cuántas consultas realiza a sus compañeros por mes?	11,41	8,39

ANOVA

	Conglomerado		Error		F	Sig.
	Media cuadrática	gl	Media cuadrática	gl		
¿Cuántas veces a la semana ingresa a la plataforma virtual de su universidad?	62,555	1	8,243	449	7,589	,006
¿Aproximadamente cuántos recursos educativos descarga de la plataforma virtual cada mes?	657,310	1	28,866	449	22,771	,000
¿Aproximadamente cuántas horas chatea sobre temas académicos cada mes?	874,842	1	30,169	449	28,998	,000
¿Aproximadamente cuántos videos académicos mira en youtube cada mes?	897,638	1	25,299	449	35,482	,000
¿Aproximadamente cuántas horas busca información académica en internet cada mes?	55112,58	2	68,580	449	803,62	,000
¿Aproximadamente cuántas consultas realiza a sus compañeros por mes?	747,509	1	58,699	449	12,735	,000

Número de casos en cada

conglomerado

Conglomerado	1	108,000
	2	343,000
Válidos		451,000
Perdidos		,000

Análisis k medias para clasificar en tres grupos a los estudiantes en función las actividades académicas.

Centros de los conglomerados finales			
	Conglomerado		
	1	2	3
¿Aproximadamente cuántos recursos educativos descarga de la plataforma virtual cada mes?	8,61	8,37	4,48
¿Aproximadamente cuántas consultas realiza a sus compañeros por mes?	11,00	12,08	5,73
¿Cuántas veces a la semana ingresa a la plataforma virtual de su universidad?	6,55	6,08	5,05
¿Aproximadamente cuántas horas chatea sobre temas académicos cada mes?	9,65	10,19	4,52
¿Aproximadamente cuántas horas busca información académica en internet cada mes?	49,39	22,37	8,67

ANOVA						
	Conglomerado		Error		F	Sig.
	Media cuadrática	gl	Media cuadrática	gl		
¿Aproximadamente cuántos recursos educativos descarga de la plataforma virtual cada mes?	869,582	2	26,515	448	32,795	,000
¿Aproximadamente cuántas consultas realiza a sus compañeros por mes?	2120,250	2	51,033	448	41,547	,000
¿Cuántas veces a la semana ingresa a la plataforma virtual de su universidad?	74,889	2	8,067	448	9,283	,000
¿Aproximadamente cuántas horas chatea sobre temas académicos cada mes?	1727,100	2	24,479	448	70,553	,000
¿Aproximadamente cuántas horas busca información académica en internet cada mes?	35429,135	2	33,586	448	1054,868	,000

Número de casos en cada conglomerado		
Conglomerado	1	51,000
	2	198,000
	3	202,000
Válidos		451,000
Perdidos		,000

Análisis k medias para clasificar en cuatro grupos a los estudiantes en función las actividades académicas.

Centros de los conglomerados finales				
	Conglomerado			
	1	2	3	4
¿Aproximadamente cuántos recursos educativos descarga de la plataforma virtual cada mes?	8,62	6,71	4,42	10,30
¿Aproximadamente cuántas consultas realiza a sus compañeros por mes?	11,22	6,24	5,56	22,55
¿Cuántas veces a la semana ingresa a la plataforma virtual de su universidad?	6,40	5,84	4,99	6,31
¿Aproximadamente cuántas horas chatea sobre temas académicos cada mes?	9,52	8,96	4,41	10,39
¿Aproximadamente cuántas horas busca información académica en internet cada mes?	49,66	22,52	7,01	18,72

ANOVA						
	Conglomerado		Error		F	Sig.
	Media cuadrática	gl	Media cuadrática	gl		
¿Aproximadamente cuántos recursos educativos descarga de la plataforma virtual cada mes?	651,127	3	26,095	447	24,952	,000
¿Aproximadamente cuántas consultas realiza a sus compañeros por mes?	5492,775	3	23,769	447	231,086	,000
¿Cuántas veces a la semana ingresa a la plataforma virtual de su universidad?	45,223	3	8,117	447	5,572	,001
¿Aproximadamente cuántas horas chatea sobre temas académicos cada mes?	905,054	3	26,187	447	34,561	,000
¿Aproximadamente cuántas horas busca información académica en internet cada mes?	24207,655	3	29,714	447	814,699	,000

Número de casos en cada conglomerado	
Conglomerado	1
	2
	3
	4
Válidos	451,000
Perdidos	,000

Análisis k medias para clasificar en cinco grupos a los estudiantes en función las actividades académicas.

Centros de los conglomerados finales					
	Conglomerado				
	1	2	3	4	5
¿Aproximadamente cuántos recursos educativos descarga de la plataforma virtual cada mes?	8,32	8,60	5,99	4,30	10,48
¿Aproximadamente cuántas consultas realiza a sus compañeros por mes?	11,12	9,96	5,51	5,61	22,95
¿Cuántas veces a la semana ingresa a la plataforma virtual de su universidad?	6,50	6,35	5,34	5,16	6,35
¿Aproximadamente cuántas horas chatea sobre temas académicos cada mes?	9,41	10,08	8,47	3,86	10,65
¿Aproximadamente cuántas horas busca información académica en internet cada mes?	54,21	31,96	18,74	6,37	17,49

ANOVA						
	Conglomerado		Error		F	Sig.
	Media cuadrática	gl	Media cuadrática	gl		
¿Aproximadamente cuántos recursos educativos descarga de la plataforma virtual cada mes?	541,069	4	25,681	446	21,069	,000
¿Aproximadamente cuántas consultas realiza a sus compañeros por mes?	3951,858	4	25,327	446	156,034	,000
¿Cuántas veces a la semana ingresa a la plataforma virtual de su universidad?	34,751	4	8,127	446	4,276	,002
¿Aproximadamente cuántas horas chatea sobre temas académicos cada mes?	827,050	4	24,916	446	33,193	,000
¿Aproximadamente cuántas horas busca información académica en internet cada mes?	19487,088	4	17,840	446	1092,331	,000

Número de casos en cada conglomerado		
Conglomerado	1	34,000
	2	72,000
	3	134,000
	4	148,000
	5	63,000
Válidos		451,000
Perdidos		,000

Anexo 3

Perfil de estudiantes en función las actividades académicas Análisis discriminante para comprobar el porcentaje de exactitud en tres grupos a los estudiantes en función las actividades académicas.

Resultados de la clasificación						
		Número inicial de casos	Grupo de pertenencia pronosticado			Total
			1	2	3	
Original	Recuento	1	50	1	0	51
		2	0	195	3	198
		3	0	9	193	202
	%	1	98,0	2,0	,0	100,0
		2	,0	98,5	1,5	100,0
		3	,0	4,5	95,5	100,0

Clasificados correctamente el 97,1% de los casos agrupados originales.

Análisis discriminante para comprobar el porcentaje de exactitud en cuatro grupos a los estudiantes en función las actividades académicas

Resultados de la clasificación							
		Número inicial de casos	Grupo de pertenencia pronosticado				Total
			1	2	3	4	
Original	Recuento	1	49	0	0	1	50
		2	0	165	0	1	166
		3	0	2	161	1	164
		4	0	0	0	71	71
	%	1	98,0	,0	,0	2,0	100,0
		2	,0	99,4	,0	,6	100,0
		3	,0	1,2	98,2	,6	100,0
		4	,0	,0	,0	100,0	100,0

a. Clasificados correctamente el 98,9% de los casos agrupados originales.

Análisis discriminante para comprobar el porcentaje de exactitud en cinco grupos a los estudiantes en función las actividades académicas

Resultados de la clasificación								
		Número inicial de casos	Grupo de pertenencia pronosticado					Total
			1	2	3	4	5	
Original	Recuento	1	34	0	0	0	0	34
		2	0	68	4	0	0	72
		3	0	0	128	5	1	134
		4	0	0	0	148	0	148
		5	0	0	0	2	61	63
	%	1	100,0	,0	,0	,0	,0	100,0
		2	,0	94,4	5,6	,0	,0	100,0
		3	,0	,0	95,5	3,7	,7	100,0
		4	,0	,0	,0	100,0	,0	100,0
		5	,0	,0	,0	3,2	96,8	100,0

Clasificados correctamente el 97,3% de los casos agrupados originales.

Anexo 4

Perfil de estudiantes en función las actividades de entretenimiento

Análisis k medias para clasificar en dos grupos a los estudiantes en función las actividades de entretenimiento

Centros de los conglomerados finales

	Conglomerado	
	1	2
¿ Aproximadamente cuántas horas a la semana chatea por diversión?	17,11	6,30
¿Aproximadamente cuántas horas a la semana utiliza redes sociales?	21,95	7,37
¿Aproximadamente cuántos videos para entretenimiento mira en YouTube cada semana?	11,84	5,87

ANOVA

	Conglomerado		Error		F	Sig.
	Media cuadrática	gl	Media cuadrática	gl		
¿ Aproximadamente cuántas horas a la semana chatea por diversión?	10951,511	1	25,856	449	423,565	,000
¿Aproximadamente cuántas horas a la semana utiliza redes sociales?	19945,489	1	34,757	449	573,853	,000
¿Aproximadamente cuántos videos para entretenimiento mira en YouTube cada semana?	3347,025	1	42,508	449	78,739	,000

Número de casos en cada conglomerado

Conglomerado	1	133,000
	2	318,000
Válidos		451,000
Perdidos		,000

Análisis k medias para clasificar en tres grupos a los estudiantes en función las actividades de entretenimiento

Centros de los conglomerados finales			
	Conglomerado		
	1	2	3
¿ Aproximadamente cuántas horas a la semana chatea por diversión?	17,77	6,32	10,42
¿Aproximadamente cuántas horas a la semana utiliza redes sociales?	23,69	7,29	11,85
¿Aproximadamente cuántos videos para entretenimiento mira en youtube cada semana?	8,53	4,65	22,62

	Conglomerado		Error		F	Sig.
	Media cuadrática	gl	Media cuadrática	gl		
¿ Aproximadamente cuántas horas a la semana chatea por diversión?	5130,777	2	27,453	448	186,891	,000
¿Aproximadamente cuántas horas a la semana utiliza redes sociales?	10462,590	2	32,648	448	320,468	,000
¿Aproximadamente cuántos videos para entretenimiento mira en youtube cada semana?	7186,677	2	17,991	448	399,467	,000

Número de casos en cada conglomerado

Conglomerado	1	106,000
	2	293,000
	3	52,000
Válidos		451,000
Perdidos		,000

Análisis k medias para clasificar en cuatro grupos a los estudiantes en función las actividades de entretenimiento

Centros de los conglomerados finales

	Conglomerado			
	1	2	3	4
¿ Aproximadamente cuántas horas a la semana chatea por diversión?	15,08	19,08	5,40	12,14
¿Aproximadamente cuántas horas a la semana utiliza redes sociales?	15,88	33,54	6,51	13,18
¿Aproximadamente cuántos videos para entretenimiento mira en YouTube cada semana?	6,56	9,97	4,78	23,20

ANOVA

	Conglomerado		Error		F	Sig.
	Media cuadrática	gl	Media cuadrática	gl		
¿ Aproximadamente cuántas horas a la semana chatea por diversión?	3819,135	3	24,839	447	153,753	,000
¿Aproximadamente cuántas horas a la semana utiliza redes sociales?	9175,451	3	17,953	447	511,076	,000
¿Aproximadamente cuántos videos para entretenimiento mira en YouTube cada semana?	4781,412	3	18,096	447	264,224	,000

Número de casos en cada conglomerado

Conglomerado	1	101,000
	2	39,000
	3	262,000
	4	49,000
Válidos		451,000
Perdidos		,000

Análisis k medias para clasificar en cinco grupos a los estudiantes en función las actividades de entretenimiento

Centros de los conglomerados finales

	Conglomerado				
	1	2	3	4	5
¿ Aproximadamente cuántas horas a la semana chatea por diversión?	15,08	19,08	19,75	7,49	5,46
¿Aproximadamente cuántas horas a la semana utiliza redes sociales?	15,88	33,54	17,81	10,05	6,52
¿Aproximadamente cuántos videos para entretenimiento mira en YouTube cada semana?	6,56	9,97	26,06	20,76	4,41

ANOVA

	Conglomerado		Error		F	Sig.
	Media cuadrática	gl	Media cuadrática	gl		
¿ Aproximadamente cuántas horas a la semana chatea por diversión?	3176,024	4	22,100	446	143,712	,000
¿Aproximadamente cuántas horas a la semana utiliza redes sociales?	6975,347	4	17,153	446	406,665	,000
¿Aproximadamente cuántos videos para entretenimiento mira en YouTube cada semana?	3864,316	4	15,641	446	247,063	,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	101,000
2	39,000
3	16,000
4	41,000
5	254,000
Válidos	451,000
Perdidos	,000

Anexo 5

Perfil de estudiantes en función las actividades de entretenimiento Análisis discriminante para comprobar el porcentaje de exactitud en tres grupos a los estudiantes en función a las actividades de entreteniendo.

Resultados de la clasificación

	Número inicial de casos	Grupo de pertenencia pronosticado			Total	
		1	2	3		
Original	Recuento	1	102	1	3	106
		2	3	289	1	293
		3	0	0	52	52
	%	1	96,2	,9	2,8	100,0
		2	1,0	98,6	,3	100,0
		3	,0	,0	100,0	100,0

Clasificados correctamente el 98,9% de los casos agrupados originales.

Análisis discriminante para comprobar el porcentaje de exactitud en cuatro grupos a los estudiantes en función a las actividades de entreteniendo

Resultados de la clasificación

	Número inicial de casos	Grupo de pertenencia pronosticado				Total	
		1	2	3	4		
Original	Recuento	1	99	0	2	0	101
		2	0	39	0	0	39
		3	6	0	252	4	262
		4	0	0	0	49	49
	%	1	98,0	,0	2,0	,0	100,0
		2	,0	100,0	,0	,0	100,0
		3	2,3	,0	96,2	1,5	100,0
		4	,0	,0	,0	100,0	100,0

Análisis discriminante para comprobar el porcentaje de exactitud en cinco grupos a los estudiantes en función a las actividades de entreteniendo

Resultados de la clasificación

	Número inicial de casos	Grupo de pertenencia pronosticado					Total	
		1	2	3	4	5		
Original	Recuento	1	100	0	0	0	1	101
		2	0	39	0	0	0	39
		3	0	0	16	0	0	16
		4	0	0	0	40	1	41
		5	3	0	0	0	251	254
	%	1	99,0	,0	,0	,0	1,0	100,0
		2	,0	100,0	,0	,0	,0	100,0
		3	,0	,0	100,0	,0	,0	100,0
		4	,0	,0	,0	97,6	2,4	100,0
		5	1,2	,0	,0	,0	98,8	100,0

Clasificados correctamente el 98,9% de los casos agrupados originales.

Anexo 6

Perfil de estudiantes en función al uso de dispositivos

Análisis k medias para clasificar en dos grupos a los estudiantes en función al uso de dispositivos

Centros de los conglomerados finales

	Conglomerado	
	1	2
Smartphone con cámara fotográfica y acceso a internet	2,41	8,62
Teléfono móvil con acceso a internet	3,55	8,80
Computador portátil	7,93	8,60

ANOVA

	Conglomerado		Error		F	Sig.
	Media cuadrática	Gl	Media cuadrática	gl		
Smartphone con cámara fotográfica y acceso a internet	4170,911	1	3,950	449	1055,879	,000
Teléfono móvil con acceso a internet	2994,327	1	5,798	449	516,413	,000
Computador portátil	48,643	1	5,919	449	8,218	,004

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	181,000
	2	270,000
Válidos		451,000
Perdidos		,000

Análisis k medias para clasificar en tres grupos a los estudiantes en función al uso de dispositivos

Centros de los conglomerados finales

	Conglomerado		
	1	2	3
Smartphone con cámara fotográfica y acceso a internet	8,45	4,20	2,80
Teléfono móvil con acceso a internet	9,16	6,22	2,21
Computador portátil	9,16	4,20	9,20

ANOVA

	Conglomerado		Error		F	Sig.
	Media cuadrática	gl	Media cuadrática	gl		
Smartphone con cámara fotográfica y acceso a internet	1517,333	2	6,495	448	233,606	,000
Teléfono móvil con acceso a internet	2046,875	2	3,357	448	609,698	,000
Computador portátil	782,382	2	2,548	448	307,010	,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	247,000
	2	76,000
	3	128,000
Válidos		451,000
Perdidos		,000

Análisis k medias para clasificar en cuatro grupos a los estudiantes en función al uso de dispositivos

Centros de los conglomerados finales

	Conglomerado			
	1	2	3	4
Smartphone con cámara fotográfica y acceso a internet	2,93	9,03	2,62	5,33
Teléfono móvil con acceso a internet	7,97	9,08	1,57	4,81
Computador portátil	8,27	9,06	9,21	2,44

ANOVA

	Conglomerado		Error		F	Sig.
	Media cuadrática	gl	Media cuadrática	gl		
Smartphone con cámara fotográfica y acceso a internet	1339,405	3	4,309	447	310,806	,000
Teléfono móvil con acceso a internet	1463,838	3	2,699	447	542,449	,000
Computador portátil	565,314	3	2,261	447	250,069	,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	75,000
	2	225,000
	3	108,000
	4	43,000
Válidos		451,000
Perdidos		,000

Anexo 7

Perfil de estudiantes en función al uso de dispositivos

Análisis discriminante para comprobar el porcentaje de exactitud en dos grupos a los estudiantes en función al uso de dispositivos

Resultados de la clasificación^a

Número inicial de casos			Grupo de pertenencia pronosticado		Total
			1	2	
Original	Recuento	1	171	10	181
		2	2	268	270
	%	1	94,5	5,5	100,0
		2	,7	99,3	100,0

a. Clasificados correctamente el 97,3% de los casos agrupados originales.

Análisis discriminante para comprobar el porcentaje de exactitud en tres grupos a los estudiantes en función al uso de dispositivos

Resultados de la clasificación^a

Número inicial de casos		Grupo de pertenencia pronosticado			Total	
		1	2	3		
Original	Recuento	1	244	1	2	247
		2	7	68	1	76
		3	3	2	123	128
	%	1	98,8	,4	,8	100,0
		2	9,2	89,5	1,3	100,0
		3	2,3	1,6	96,1	100,0

Clasificados correctamente el 96,5% de los casos agrupados originales.

Análisis discriminante para comprobar el porcentaje de exactitud en cuatro grupos a los estudiantes en función al uso de dispositivos

Resultados de la clasificación^a

Número inicial de casos			Grupo de pertenencia pronosticado				Total
			1	2	3	4	
Original	Recuento	1	71	1	1	2	75
		2	0	221	2	2	225
		3	6	0	102	0	108
		4	0	0	1	42	43
	%	1	94,7	1,3	1,3	2,7	100,0
		2	,0	98,2	,9	,9	100,0
		3	5,6	,0	94,4	,0	100,0
		4	,0	,0	2,3	97,7	100,0

a. Clasificados correctamente el 96,7% de los casos agrupados originales.

Anexo 8

Pruebas de chi cuadrado para establecer Relación entre perfiles de uso de Internet en actividades académicas e Ingresos

Pruebas de chi-cuadrado

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	15,999 ^a	8	,042
Razón de verosimilitudes	16,668	8	,034
Asociación lineal por lineal	9,634	1	,002
N de casos válidos	451		

a. 0 casillas (,0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 7,69.

Pruebas de chi cuadrado para establecer Relación entre perfiles de uso de Internet en actividades de entretenimiento e Ingresos

Pruebas de chi-cuadrado

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	9,522 ^a	8	,300
Razón de verosimilitudes	9,632	8	,292
Asociación lineal por lineal	,035	1	,852
N de casos válidos	451		

a. 0 casillas (,0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 7,84.

Anexo 9

Pruebas de chi cuadrado para establecer Relación entre perfiles de uso de Internet en actividades académicas y Rendimiento académico

Pruebas de chi-cuadrado

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	9,690 ^a	6	,138
Razón de verosimilitudes	10,030	6	,123
Asociación lineal por lineal	,777	1	,378
N de casos válidos	451		

a. 4 casillas (33,3%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 1,13.

Pruebas de chi cuadrado para establecer Relación entre perfiles de uso de Internet en actividades de entretenimiento y Rendimiento académico

Pruebas de chi-cuadrado

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	4,445 ^a	6	,617
Razón de verosimilitudes	4,092	6	,664
Asociación lineal por lineal	,006	1	,939
N de casos válidos	451		

a. 3 casillas (25,0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 1,15.

Anexo 10

Regresión Logística Multinomial Hipótesis 1

Codificación de la variable dependiente

Valor original	Valor interno
Moderno	0
Tradicional	1

Bloque 1: Método = Introducir

Pruebas omnibus sobre los coeficientes del modelo

		Chi cuadrado	gl	Sig.
Paso 1	Paso	10,637	4	,031
	Bloque	10,637	4	,031
	Modelo	10,637	4	,031

Resumen del modelo

Paso	-2 log de la verosimilitud	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	485,883 ^a	,023	,035

a. La estimación ha finalizado en el número de iteración 4 porque las estimaciones de los parámetros han cambiado en menos de ,001.

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

		Número inicial de casos = Moderno		Número inicial de casos = Tradicional		Total
		Observado	Esperado	Observado	Esperado	
Paso 1	1	25	25,000	43	43,000	68
	2	20	20,000	56	56,000	76
	3	18	18,000	58	58,000	76
	4	28	28,000	93	93,000	121
	5	17	17,000	93	93,000	110

Tabla de clasificación

Observado			Pronosticado		
			Número inicial de casos		Porcentaje correcto
			Moderno	Tradicional	
Paso 1	Número inicial de casos	Moderno	0	108	,0
		Tradicional	0	343	100,0
	Porcentaje global				76,1

a. El valor de corte es ,500

Variables en la ecuación

	B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95% para EXP(B)	
							Inferior	Superior
Paso 1 ^a ing			10,393	4	,034			
ing(1)	,628	,369	2,896	1	,089	1,873	,909	3,860
ing(2)	1,157	,364	10,079	1	,001	3,181	1,557	6,497
ing(3)	,658	,331	3,947	1	,047	1,931	1,009	3,696
ing(4)	,487	,362	1,811	1	,178	1,628	,801	3,310
Constante	,542	,252	4,650	1	,031	1,720		

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

Anexo 11

Regresión Logística Multinomial Hipótesis 2

Codificación de la variable dependiente

Valor original	Valor interno
Modernos	0
Pasivos	1

Tabla de clasificación^{a,b}

Observado			Pronosticado		
			Número inicial de casos		Porcentaje correcto
			Modernos	Pasivos	
Paso 0	Número inicial de casos	Modernos	0	133	,0
		Pasivos	0	318	100,0
	Porcentaje global				70,5

a. En el modelo se incluye una constante.

b. El valor de corte es ,500

Bloque 1: Método = Introducir

Pruebas omnibus sobre los coeficientes del modelo

		Chi cuadrado	Gl	Sig.
Paso 1	Paso	4,401	4	,355
	Bloque	4,401	4	,355
	Modelo	4,401	4	,355

Resumen del modelo

Paso	-2 log de la verosimilitud	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	542,645 ^a	,010	,014

Resumen del modelo

Paso	-2 log de la verosimilitud	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	542,645 ^a	,010	,014

a. La estimación ha finalizado en el número de iteración 4 porque las estimaciones de los parámetros han cambiado en menos de ,001.

Prueba de Hosmer y Lemeshow

Paso	Chi cuadrado	gl	Sig.
1	,000	3	1,000

Tabla de clasificación^a

Observado			Pronosticado		
			Número inicial de casos		Porcentaje correcto
			Modernos	Pasivos	
Paso 1	Número inicial de casos	Modernos	0	133	,0
		Pasivos	0	318	100,0
	Porcentaje global				70,5

a. El valor de corte es ,500

Variables en la ecuación

	B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95% para EXP(B)	
							Inferior	Superior
Paso 1 ^a ing			4,453	4	,348			
ing(1)	,487	,362	1,811	1	,178	1,628	,801	3,310
ing(2)	,485	,332	2,135	1	,144	1,624	,847	3,112
ing(3)	,481	,325	2,184	1	,139	1,617	,855	3,058
ing(4)	,054	,347	,024	1	,877	1,055	,534	2,085
Constante	,542	,252	4,650	1	,031	1,720		

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

Anexo 12

Regresión Logística Multinomial Hipótesis 3

Codificación de la variable

dependiente

Valor original	Valor interno
Aprobado	0
Reprobad	1

Bloque 0: Bloque inicial

Tabla de clasificación^{a,b}

Observado			Pronosticado		
			rendimiento_acad		Porcentaje correcto
			aprobado	reprobad	
Paso 0	rendimiento_acad	Aprobado	349	0	100,0
		Reprobad	102	0	,0
		Porcentaje global			77,4

a. En el modelo se incluye una constante.

b. El valor de corte es ,500

Bloque 1: Método = Introducir

Pruebas omnibus sobre los coeficientes del modelo

		Chi cuadrado	Gl	Sig.
Paso 1	Paso	,171	1	,680
	Bloque	,171	1	,680
	Modelo	,171	1	,680

Resumen del modelo

Paso	-2 log de la verosimilitud	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	482,038 ^a	,000	,001

Resumen del modelo

Paso	-2 log de la verosimilitud	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	482,038 ^a	,000	,001

a. La estimación ha finalizado en el número de iteración 4 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	.

Variables en la ecuación

Paso	Cluster_acade_2 (1)	B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95% para EXP(B)	
								Inferior	Superior
	Constante	-1,257	,130	93,404	1	,000	,285		
		,108	,260	,172	1	,678	1,114	,669	1,854

a. Variable(s) introducida(s) en el paso 1: Cluster_acade_2.

Anexo 13

Regresión Logística Multinomial Hipótesis 4

Codificación de la variable dependiente

Valor original	Valor interno
Aprobado	0
Reprobad	1

Bloque 0: Bloque inicial

Tabla de clasificación^{a,b}

Observado			Pronosticado		
			rendimiento_acad		Porcentaje correcto
			aprobado	reprobad	
Paso 0	rendimiento_acad	Aprobado	349	0	100,0
		Reprobad	102	0	,0
		Porcentaje global			77,4

a. En el modelo se incluye una constante.

b. El valor de corte es ,500

Bloque 1: Método = Introducir

Pruebas omnibus sobre los coeficientes del modelo

		Chi cuadrado	gl	Sig.
Paso 1	Paso	2,085	1	,149
	Bloque	2,085	1	,149
	Modelo	2,085	1	,149

Resumen del modelo

Paso	-2 log de la verosimilitud	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	480,124 ^a	,005	,007

a. La estimación ha finalizado en el número de iteración 4 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	.

Variables en la ecuación

	B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95% para EXP(B)	
							Inferior	Superior
Paso 1 ^a clus_entre2(1)	,349	,239	2,124	1	,145	1,417	,887	2,265
Constante	-1,340	,138	93,882	1	,000	,262		

a. Variable(s) introducida(s) en el paso 1: clus_entre2.

Anexo 14

Base de datos realizando la codificación de la nueva variable uso_internet_academico en Excel.

Umbrales_medidas - Microsoft Excel

Archivo Inicio Insertar Diseño de página Fórmulas Datos Revisar Vista

Calibri 11 Fuente Ajustar texto General

Formato condicional Dar formato como tabla Estilos de celda Insertar Eliminar Formato Celdas

Autosuma Rellenar Ordenar y filtrar Buscar y seleccionar

Portapapeles Copiar formato Copiar formato Alineación Número

Q25 =SI(L25:L475=1;"MENOR";"MAYOR")

	I	J	K	L	M	N	O	P	Q
1	a_hor_ci	a_bus_i	a_bib_v	uso_internet_academico_6	uso_internet_academico_5	uso_internet_academico_4	uso_internet_academico_3	uso_internet_academico_2	uso_internet_academico_1
2	1	20	0	2	2	2	1	2	2 MAYOR
3	4	30	2	2	2	2	1	2	2 MAYOR
4	15	10	0	2	2	1	1	1	1 MAYOR
5	10	15	8	1	1	2	1	1	1 MENOR
6	5	20	10	2	2	1	1	1	2 MAYOR
7	5	5	0	2	2	1	2	2	2 MAYOR
8	10	20	30	2	2	2	1	1	2 MAYOR
9	5	20	0	1	2	1	1	1	1 MENOR
10	6	50	4	2	2	2	2	2	2 MAYOR
11	10	60	15	2	2	2	2	2	2 MAYOR
12	2	8	5	2	2	2	2	2	2 MAYOR
13	2	6	2	2	2	2	2	2	2 MAYOR
14	0	40	0	2	2	2	1	1	1 MAYOR
15	0	0	0	2	2	2	1	1	2 MAYOR
16	0	20	20	2	2	1	1	1	2 MAYOR
17	1	50	10	2	2	2	2	2	2 MAYOR
18	4	19	20	2	2	2	2	2	2 MAYOR
19	4	10	2	1	2	1	1	1	2 MENOR
20	6	7	8	2	2	2	1	1	2 MAYOR
21	2	20	0	2	2	2	1	1	1 MAYOR
22	10	20	0	2	2	2	1	1	2 MAYOR
23	10	10	20	2	2	2	1	1	2 MAYOR
24	5	10	0	2	2	2	1	1	1 MAYOR
25	10	15	0	2	2	1	1	1	1 MAYOR

Anexo 15

Selección de variables de uso de internet académico

Hipótesis 1: El nivel de ingresos determina como se utiliza internet para el aprendizaje.

Attribute selection

CfsSubsetEval

ScatterSearchV1

EXPERIMENTO 1

Attribute Evaluator	CfsSubsetEval		
Search Method	GreedyStepwise		
Cross-validation	Folds	5	
	Seed	1	
#	Attribute	# Folds	%
1	a_ing_pla	2	40
2	a_con_pro	3	60
3	a_con_est	3	60
4	a_rec_edu	1	20
5	a_vid_aca	2	40
6	a_for_vir	4	80
7	a_post_aca	5	100
8	a_hor_cha	0	0
9	a_bus_inf	5	100

EXPERIMENTO 2

Attribute Evaluator	CfsSubsetEval		
Search Method	GreedyStepwise		
Cross-validation	Folds	10	
	Seed	1	
#	Attribute	# Folds	%
1	a_ing_pla	6	60
2	a_con_pro	9	90
3	a_con_est	8	80
4	a_rec_edu	4	40
5	a_vid_aca	7	70
6	a_for_vir	9	90
7	a_post_aca	10	100
8	a_hor_cha	1	10
9	a_bus_in	10	100

EXPERIMENTO 3

Attribute Evaluator	CfsSubsetEval		
Search Method	GreedyStepwise		
Cross-validation	Folds	15	
	Seed	1	
#	Attribute	# Folds	%
1	a_ing_pla	10	67
2	a_con_pro	13	87
3	a_con_est	10	67
4	a_rec_edu	6	40
5	a_vid_aca	10	67
6	a_for_vir	14	93
7	a_post_aca	15	100
8	a_hor_cha	0	0
9	a_bus_inf	15	100

EXPERIMENTO 4

Attribute Evaluator	CfsSubsetEval		
Search Method	GreedyStepwise		
Cross-validation	Folds	20	
	Seed	1	
#	Attribute	# Folds	%
1	a_ing_pla	17	85
2	a_con_pro	20	100
3	a_con_est	13	65
4	a_rec_edu	6	30
5	a_vid_aca	13	65
6	a_for_vir	20	100
7	a_post_aca	20	100
8	a_hor_cha	0	0
9	a_bus_inf	20	100

EXPERIMENTO 5

Attribute Evaluator	CfsSubsetEval		
Search Method	GreedyStepwise		
Cross-validation	Folds	25	
	Seed	1	
#	Attribute	# Folds	%
1	a_ing_pla	20	80
2	a_con_pro	25	100
3	a_con_est	21	84
5	a_vid_aca	20	80
6	a_for_vir	25	100
7	a_post_aca	25	100
9	a_bus_inf	25	100

Resultados

Attribute	EXPERIMENTO 1	EXPERIMENTO 2	EXPERIMENTO 3	EXPERIMENTO 4	EXPERIMENTO 5
	%	%	%	%	%
a_con_pro	60	90	87	100	100
a_con_est	60	80	67	65	84
a_for_vir	80	90	93	100	100
a_post_aca	100	100	100	100	100
a_bus_inf	100	100	100	100	100
a_ing_pla	N/A	60	67	85	80
a_vid_aca	N/A	70	67	65	80

Selección de variables de uso de internet para entretenimiento y diversión

- **Hipótesis 2:** El nivel de ingresos económicos determina como se utiliza internet para entretenimiento

EXPERIMENTO 1

Attribute Evaluator	CfsSubsetEval		
Search Method	GreedyStepwise		
Cross-validation	Folds	5	
	Seed	1	
# Folds	%	Atributo	
5	100	2	e_hor_cha
0	0	3	e_red_soc
1	20	4	e_jue_lin
4	80	5	e_des_con
4	80	6	e_vid_sem

EXPERIMENTO 2

Attribute Evaluator	CfsSubsetEval		
Search Method	GreedyStepwise		
Cross-validation	Folds	10	
	Seed	1	
# Folds	%	Atributo	
10	100	2	e_hor_cha
0	0	3	e_red_soc
1	10	4	e_jue_lin
10	100	5	e_des_con
8	80	6	e_vid_sem

EXPERIMENTO 3

Attribute Evaluator	CfsSubsetEval		
Search Method	GreedyStepwise		
Cross-validation	Folds	15	
	Seed	1	
# Folds	%	Atributo	
15	100	2	e_hor_cha
0	0	3	e_red_soc
1	7	4	e_jue_lin
15	100	5	e_des_con
9	60	6	e_vid_sem

EXPERIMENTO 4

Attribute Evaluator	CfsSubsetEval		
Search Method	GreedyStepwise		
Cross-validation	Folds	20	
	Seed	1	
# Folds	%	Atributo	
20	100	2	e_hor_cha
0	0	3	e_red_soc
1	5	4	e_jue_lin
20	100	5	e_des_con
14	70	6	e_vid_sem

EXPERIMENTO 5

Attribute Evaluator	CfsSubsetEval		
Search Method	GreedyStepwise		
Cross-validation	Folds	25	
	Seed	1	
# Folds	%	Atributo	
25	100	2	e_hor_cha
0	0	3	e_red_soc
1	4	4	e_jue_lin
25	100	5	e_des_con
14	56	6	e_vid_sem

Attribute	EXPERIMENTO	EXPERIMENTO	EXPERIMENTO	EXPERIMENTO	EXPERIMENTO
	1	2	3	4	5
	%	%	%	%	%
e_hor_cha	100	100	100	100	100
e_red_soc	0	0	0	0	0
e_jue_lin	20	10	7	5	4
e_des_con	80	100	100	100	100
e_vid_sem	80	80	60	70	56

Anexo 16

Experimentos con Umbrales de la variable niv_con

Umbral 6

UMBRAL6	1	1'6	2	7'10
ERROR	123,001272	11,0905939		
cluster	full	Cluster 0	Cluster 1	
instances	451	363	88	
ing	3	3	3	
a_for_vir	2	2	2	
a_post_aca	3	3	2	
a_bus_inf	19	19	19	
uso_internet_academico_6	2	2	1	

Umbral 5

EXPERIMENTO 2				
UMBRAL 5	1	1'5	2	6'10
ERROR	92,5196056	9,61871122		
cluster	full	Cluster 0	Cluster 1	
instances	297	88	209	
ing	3	3	3	
a_for_vir	2	4	1	
a_post_aca	3	4	2	
a_bus_inf	19	20	19	
uso_internet_academico_5	2	2	2	

Umbral 4

UMBRAL 5	1	1'4	2	5'10
ERROR	172,519606	13,1346719		
cluster	full	Cluster 0	Cluster 1	
instances	297	88	209	
ing	3	3	3	
a_for_vir	2	4	1	
a_post_aca	3	4	2	
a_bus_inf	19	20	19	
uso_internet_academico_4	2	2	2	

Umbral 3

EXPERIMENTO 4				
UMBRAL 4	1	1'3	2	4'10
ERROR	80,7044616	8,9835662		
cluster	full	Cluster 0	Cluster 1	
instantes	297	131	166	
ing	3	3	3	
a_for_vir	2	2	2	
a_post_aca	3	3	2	
a_bus_inf	19	21	18	
uso_internet_academico_3	1	2	1	

Umbral 2

EXPERIMENTO 5				
UMBRAL 5	1	1'2	2	3'10
ERROR	119,519606	10,9325023		
cluster	full	Cluster 0	Cluster 1	
instances	297	88	209	
ing	3	3	3	
a_for_vir	2	4	1	
a_post_aca	3	4	2	
a_bus_inf	19	20	19	
uso_internet_academico_2	2	2	2	

Resultado mejor Umbral

EXPERIMENTO	sum of squared errors	RAÍZ
1	123,001272	11,0905939
2	92,5196056	9,61871122
3	172,519606	13,1346719
4	80,7044616	8,9835662
5	119,519606	10,9325023

Anexo 17

Resultados Obtenidos para las 4 hipótesis con varios experimentos (Hipótesis 1)

Seed 10

EXPERIMENTO 1			
sum of squared errors	260,490064	Raíz	16,1397046
Seed	10		
Cluster#			
Attribute	Full Data	0	1
	451	193	258
ing	2,8891	4,0984	1,9845
a_for_vir	2,0222	1,601	2,3372
a_post_aca	2,6142	2,4456	2,7403
a_bus_inf	19,2905	22,8912	16,5969
uso_internet_academico_4	2	2	2
UMBRAL 4		1 (1/3)	2 (4/10)

Seed 15

EXPERIMENTO 2			
sum of squared errors	269,88222	Raíz	16,4280924
Seed	15		
Cluster#			
Attribute	Full Data	0	1
	451	133	318
ing	2,8891	2,8571	2,9025
a_for_vir	2,0222	3,8045	1,2767
a_post_aca	2,6142	4,015	2,0283
a_bus_inf	19,2905	21,0977	18,5346
uso_internet_academico_4	2	2	2
UMBRAL 4		1 (1/3)	2 (4/10)

Seed 20

EXPERIMENTO 3			
sum of squared errors	121,677974	Raíz	11,0307739
Seed	20		
Cluster#			
Attribute	Full Data	0	1
	451	171	280
ing	2,8891	2,7018	3,0036
a_for_vir	2,0222	1,7368	2,1964
a_post_aca	2,6142	2,4971	2,6857
a_bus_inf	19,2905	16,117	21,2286
uso_internet_academico_4	2	1	2

UMBRAL 4		1 (1/3)	2 (4/10)
----------	--	---------	----------

Seed 25

EXPERIMENTO 4			
sum of squared errors	121,677974	Raíz	11,0307739
Seed	25		
Cluster#			
Attribute	Full Data	0	1
	451	171	280
ing	2,8891	2,7018	3,0036
a_for_vir	2,0222	1,7368	2,1964
a_post_aca	2,6142	2,4971	2,6857
a_bus_inf	19,2905	16,117	21,2286
uso_internet_academico_4	2	1	2
UMBRAL 4		1 (1/3)	2 (4/10)

Seed 30

EXPERIMENTO 5			
sum of squared errors	221,677974	Raíz	14,888854
Seed	30		
Cluster#			
Attribute	Full Data	0	1
	451	280	171
ing	2,8891	3,0036	2,7018
a_for_vir	2,0222	2,1964	1,7368
a_post_aca	2,6142	2,6857	2,4971
a_bus_inf	19,2905	21,2286	16,117
uso_internet_academico_4	2	2	1
UMBRAL 4		1 (1/3)	2 (4/10)

Resultado mejor Seed

EXPERIMENTO	SEED	sum of squared errors	RAÍZ
1	10	260,490064	16,1397046
2	15	269,88222	16,4280924
3	20	121,677974	11,0307739
4	25	121,677974	11,0307739
5	30	221,677974	14,888854

✓ Hipótesis 2

EXPERIMENTO 1			
sum of squared errors	119,7894	Raíz	10,9448365
Seed	10		
Cluster#			
Attribute	Full Data	0	1
	451	245	206
ing	2,8891	2,1592	3,7573
e_hor_cha	9,4856	6,1265	13,4806
e_red_soc	11,6718	8,3102	15,6699
e_jue_lin	2,6829	1,6	3,9709
e_des_con	4,0244	2,7673	5,5194
e_vid_sem	7,6297	4,751	11,0534

EXPERIMENTO 2			
sum of squared errors	119,7894	Raíz	10,9448365
Seed	15		
Cluster#			
Attribute	Full Data	0	1
	451	245	206
ing	2,8891	2,1592	3,7573
e_hor_cha	9,4856	6,1265	13,4806
e_red_soc	11,6718	8,3102	15,6699
e_jue_lin	2,6829	1,6	3,9709
e_des_con	4,0244	2,7673	5,5194
e_vid_sem	7,6297	4,751	11,0534

EXPERIMENTO 3			
sum of squared errors	123,3592	Raíz	11,1067206
Seed	20		
Cluster#			
Attribute	Full Data	0	1
	451	264	187
ing	2,8891	3,803	1,5989
e_hor_cha	9,4856	10,1364	8,5668
e_red_soc	11,6718	12,2727	10,8235
e_jue_lin	2,6829	2,8977	2,3797
e_des_con	4,0244	4,3371	3,5829
e_vid_sem	7,6297	8,1894	6,8396
EXPERIMENTO 4			

sum of squared errors	119,7894	Raíz	10,944833
Seed	25		
Cluster#			
Attribute	Full Data	0	1
	451	241	210
Ing	2,8891	2,1452	3,7429
e_hor_cha	9,4856	6,1245	13,3429
e_red_soc	11,6718	8,39	15,4381
e_jue_lin	2,6829	1,5768	3,9524
e_des_con	4,0244	2,639	5,6143
e_vid_sem	7,6297	4,7469	10,9381

EXPERIMENTO 5			
sum of squared errors	123,3592	Raíz	11,1067206
Seed	30		
Cluster#			
Attribute	Full Data	0	1
	451	264	187
Ing	2,8891	3,803	1,5989
e_hor_cha	9,4856	10,1364	8,5668
e_red_soc	11,6718	12,2727	10,8235
e_jue_lin	2,6829	2,8977	2,3797
e_des_con	4,0244	4,3371	3,5829
e_vid_sem	7,6297	8,1894	6,8396

EXPERIMENTO	SEED	sum of squared errors	RAÍZ
1	10	119,7894	10,9448365
2	15	119,7894	10,9448365
3	20	123,3592	11,1067206
4	25	119,7894	10,944833
5	30	123,3592	11,1067206

✓ Hipótesis 3

✓ EXPERIMENTO 1			
sum of squared errors	51,5826	Raíz	7,18210324
Seed	10		
Cluster#			
Attribute	Full Data	0	1
	451	278	173
a_for_vir	2,0222	1,1799	3,3757
a_post_aca	2,6142	1,5899	4,2601
a_bus_inf	19,2905	15,6763	25,0983
		APROBADOS	REPROBADO AL MENOS 1
Incorrectly clustered instances		185,0000	41,0200

EXPERIMENTO 2			
sum of squared errors	51,6585	Raíz	7,18738313
Seed	15		
Cluster#			
Attribute	Full Data	0	1
	451	129	322
a_for_vir	2,0222	3,8605	1,2857
a_post_aca	2,6142	3,5969	2,2205
a_bus_inf	19,2905	21,0930	18,5683
		APROBADOS	REPROBADO AL MENOS 1
Incorrectly clustered instances		165,0000	36,5854

EXPERIMENTO 3			
sum of squared errors	51,5826	Raíz	7,18210324
Seed	20		
Cluster#			
Attribute	Full Data	0	1
	451	173	278
a_for_vir	2,0222	3,3757	1,1799
a_post_aca	2,6142	4,2601	1,5899
a_bus_inf	19,2905	25,0983	15,6763
		APROBADOS	REPROBADO AL MENOS 1
Incorrectly clustered instances		185,0000	41,0200

EXPERIMENTO 4			
sum of squared errors	51,5824	Raíz	7,18209044
Seed	25		
Cluster#			
Attribute	Full Data	0	1
	451	279	172
a_for_vir	2,0222	1,1828	3,3837
a_post_aca	2,6142	1,6201	4,2267
a_bus_inf	19,2905	15,6631	25,1744
		APROBADOS	REPROBADO AL MENOS 1
Incorrectly clustered instances		184,0000	40,7982

EXPERIMENTO 5			
sum of squared errors	51,6157	Raíz	7,18440941
Seed	30		
Cluster#			
Attribute	Full Data	0	1
	451	133	318
a_for_vir	2,0222	3,8045	1,2767
a_post_aca	2,6142	4,0150	2,0283
a_bus_inf	19,2905	21,0977	18,5346
		APROBADOS	REPROBADO AL MENOS 1
Incorrectly clustered instances		167,0000	37,0288

EXPERIMENTO	SEED	sum of squared errors	RAÍZ	INSTANCIAS INCORRECTAS
1	10	51,5826	7,18210324	185,0000
2	15	51,6585	7,18738313	165,0000
3	20	51,5826	7,18210324	185,0000
4	25	51,5824	7,18209044	184,0000
5	30	51,6157	7,18440941	167,0000

✓ Hipótesis 4

✓ EXPERIMENTO 1			
sum of squared errors	75,5090	Raíz	8,68959052
Seed	10		
Cluster#			
Attribute	Full Data	0	1
	451	102	349
e_hor_cha	9,4856	9,3039	9,5387
e_des_con	4,0244	3,6078	4,1461
e_vid_sem	7,6297	8,3627	7,4155
U_entre_rendimiento	20	10	20

EXPERIMENTO 2			
sum of squared errors	150,4422	Raíz	12,26549
Seed	15		
Cluster#			
Attribute	Full Data	0	1
	451	285	166
e_hor_cha	9,4856	6,8456	14,0181
e_des_con	4,0244	2,1263	7,2831
e_vid_sem	7,6297	4,4211	13,1386
U_entre_rendimiento	20,0000	20,0000	20,0000

EXPERIMENTO 3			
sum of squared errors	75,5090	Raíz	8,68959052
Seed	20		
Cluster#			
Attribute	Full Data	0	1
	451	102	349
e_hor_cha	9,4856	9,3039	9,5387
e_des_con	4,0244	3,6078	4,1461
e_vid_sem	7,6297	8,3627	7,4155
U_entre_rendimiento	20,0000	10,0000	20,0000

EXPERIMENTO 4			
sum of squared errors	150,4464	Raíz	12,265658
Seed	25		
Cluster#			
Attribute	Full Data	0	1
	451	279	172
e_hor_cha	9,4856	6,7563	13,9128
e_des_con	4,0244	2,1039	7,1395
e_vid_sem	7,6297	4,2652	13,0872
U_entre_rendimiento	20,0000	20,0000	20,0000

EXPERIMENTO 5			
sum of squared errors	150,4464	Raíz	12,265658
Seed	30		
Cluster#			
Attribute	Full Data	0	1
	451	279	172
e_hor_cha	9,4856	6,7563	13,9128
e_des_con	4,0244	2,1039	7,1395
e_vid_sem	7,6297	4,2652	13,0872
U_entre_rendimiento	20,0000	20,0000	20,0000

EXPERIMENTO	SEED	sum of squared errors	RAÍZ
1	10	75,5090	8,68959052
2	15	150,4422	12,26549
3	20	75,5090	8,68959052
4	25	150,4464	12,265658
5	30	150,4464	12,265658