UNIVERSIDAD TÉCNICA PARTICULAR DE LOJA

La Universidad Católica de Loja

ÁREA TÉCNICA

TITULO DE INGENIERO EN SISTEMAS INFORMÁTICOS Y COMPUTACIÓN

Determinación de perfiles de uso de redes sociales en los habitantes de la provincia de Loja.

TRABAJO DE TITULACIÓN.

AUTORA: Mendoza Calva, Andrea Cristina

DIRECTOR: Torres Díaz, Juan Carlos, Mgs.

LOJA – ECUADOR

2017

Septiembre, 2017

APROBACIÓN DE DIRECTOR DE TRABAJO DE TITULACIÓN

Magister.
Juan Carlos Torres Díaz
DOCENTE DE LA TITULACION

De mi consideración:

El presente trabajo de fin de titulación: *Determinación de perfiles de uso de redes sociales en los habitantes de la provincia de Loja* realizado por: *Andrea Cristina Mendoza Calva* ha sido orientado y revisado durante su ejecución, por lo que se aprueba la presentación del mismo.

Loja, Julio de 2017

f) ……

DECLARACIÓN DE AUTORÍA Y CESIÓN DE DERECHOS

“Yo Mendoza Calva Andrea Cristina declaro ser autor (a) del presente trabajo de fin de titulación: Determinación de perfiles de uso de redes sociales en los habitantes de la provincia de Loja de la Titulación de Sistemas Informáticos y Computación, siendo Juan Carlos Torres Díaz director del presente trabajo; y eximo expresamente a la Universidad Técnica Particular de Loja y a sus representantes legales de posibles reclamos o acciones legales. Además, certifico que las ideas, conceptos, procedimientos y resultados vertidos en el presente trabajo investigativo, son de mi exclusiva responsabilidad.

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

f.……

Mendoza Calva Andrea Cristina
1105024457

DEDICATORIA

A mis padres Marco y Mélida, por ser los pilares fundamentales en mi vida.

A Ricardo, Camila y Matías, por ser mi más grande inspiración y deseo de superación.

A Daniel y Nathaly que más que mis hermanos son mis amigos y quienes han estado junto a mi incondicionalmente.

Con mucho carriño.

Andrea.

AGRADECIMIENTO

El culminar con el trabajo de titulación ha supuesto un importante enriquecimiento personal y a la vez una gran satisfacción, pues lo que se obtiene con esfuerzo y dedicación es lo que más se valora. En este camino se han hecho presentes obstáculos y dificultades, sin embargo, la ayuda, apoyo y compañía de varias personas a las cuales quiero expresar mi agradecimiento.

A mi director Juan Carlos Torres, por la confianza, el apoyo, las enseñanzas y la paciencia dedicada en la orientación de este trabajo.

A la Universidad Técnica Particular de Loja por la formación académica y a todos mis profesores que desinteresadamente compartieron su conocimiento.

Finalmente, a mis familiares y amigos, por los consejos, la confianza y el apoyo, que con el pasar de los años me han otorgado.

A ustedes, eternamente agradecida.

Andrea.

INDICE DE CONTENIDO

INDICE DE TABLAS

INDICE DE ECUACIONES

INDICE DE FIGURAS

RESUMEN

El avance de la tecnología se ha incorporado en la vida diaria de los seres humanos, generando cambios y necesidades, que describen la participación de usuarios en redes sociales.

El propósito del presente trabajo de titulación es determinar el patrón de consumo digital de los habitantes de la provincia de Loja, al aplicar técnicas de minería de datos como clusterización, K means, mapas auto organizados de Kohonen, arboles de decisión dentro de la herramienta R y el análisis estadístico con cruce de variables al contenido de los perfiles tecnológicos utilizando la metodología KDD.

Presentando como resultado que las técnicas aplicadas, permiten generar contenido participativo y personalizado; por la determinación más óptima de perfiles de uso con contenido social, reflexivo, familiar y de humor de los usuarios de la provincia de Loja.

PALABRAS CLAVES:

Redes sociales, minería de datos, clusterización, k means, mapas auto organizados de Kohonen, análisis estadístico, perfiles de uso.

ABSTRACT

The advance of technology has been incorporated into the daily life of human beings, generating changes and needs, which describe the participation of users in social networks.

The purpose of the present titling work is to determine the digital consumption pattern of the inhabitants of the province of Loja, applying data mining techniques such as clustering, K means, Kohonen self-organized maps, decision trees within tool R And statistical analysis with crossing of variables to the content of the technological profiles using the KDD methodology.

As a result, the techniques applied allow the generation of participative and personalized content; By the most optimal determination of use profiles with social, reflective, familiar and humor content of the users of the province of Loja.

KEYWORDS:

Social networks, data mining, clustering, k means, Kohonen self-organized maps, statistical analysis, usage profiles.

CAPITULO I

1. INTRODUCCION

En el presente capítulo se describe la Introducción del Trabajo de Titulación destacando los objetivos y la estructura de desarrollo.

1.1 Introducción

Actualmente las redes sociales se han incorporado en la vida diaria de los seres humanos, obligando a los medios a una evolución constante, lo cual requiere de filtros que permitan que la información llegue acorde a las necesidades de cada persona para que sea agradable y de utilidad. Es por esto que el Trabajo de Titulación determina patrones de consumo digital de los habitantes de la provincia de Loja, mediante la selección de perfiles de los usuarios de más de 15 años (N=502), examinando el contenido de sus publicaciones.

Las herramientas tecnológicas se utilizan para el tratamiento y análisis de acuerdo a las fases de la metodología KDD (Knowledge Discovery in databases); gracias al uso diario de redes sociales, se permitió la selección de perfiles con la extracción del contenido para una clasificación, distribución y determinación de categorías.

La información recolectada permite establecer una relación propicia a través del contenido social, reflexivo, familiar y de humor, donde los usuarios generan patrones de uso en la red, describiendo un consumo con pautas concretas para un trabajo participativo y personalizado.

Las actividades realizadas contribuyen con un nuevo conocimiento para la producción de contenido, dando apertura a una comunicación más directa.

1.2 Objetivos

De acuerdo al análisis realizado para el desarrollo del presente Trabajo de Titulación y el uso masivo de medios digitales por la sociedad en general se han definido los siguientes objetivos.

1.2.1 Objetivo general.

Determinación de los patrones de consumo de contenidos digitales en la red social facebook para los habitantes de la provincia de Loja.

1.2.2 Objetivos específicos.

- Conocer preferencias de información y contenido.
- Identificar el interés por la información, atribuido al comportamiento y contenido de su perfil dentro de redes sociales.

1.3 Estructura del documento

El presente Trabajo de Titulación se estructura en 5 capítulos:

- El capítulo I muestra la Introducción, objetivos y estructura; aquí se resalta la importancia y aplicación del Trabajo de Titulación con el uso adecuado de redes sociales.
- El capítulo II es el marco teórico, donde se investiga temas como el uso de nuevas tecnologías, brecha digital, consumo digital, redes sociales y minería de datos.
- El capítulo III describe la metodología KDD utilizada para el desarrollo del presente Trabajo de Titulación.
- El capítulo IV presenta los resultados y el análisis obtenidos con la aplicación de algoritmos de minería de datos para determinar patrones de preferencia de contenido en los perfiles de usuarios de redes sociales.

Finalmente, se presentan las conclusiones obtenidas al finalizar el Trabajo de Titulación y los anexos correspondientes.

CAPITULO II

2. MARCO TEORICO

En el presente capítulo se describe el marco teórico considerando los fundamentos técnicos para el desarrollo del Trabajo de Titulación.

2.1 Nuevas tecnologías

La información y la forma de comunicación en la sociedad ha cambiado de una manera innegable, provocando una variación en lo económico y político.

Gracias a la accesibilidad y facilidad de uso que brindan los dispositivos electrónicos como el computador (portátil y de escritorio), Smartphone y Tablet es posible contar con información actualizada y en el momento que se precise, llegando a convirtiéndose en una necesidad de la sociedad.

En la Figura 1 y de acuerdo al Instituto nacional de estadística y censos (INEC) se puede determinar el estado tecnológico de Ecuador.

Figura 1. Estado Tecnológico del Ecuador
Fuente: Encuesta Nacional de Empleo, Desempleo y Subempleo, diciembre 2015
Elaboración: Instituto Nacional de Estadística y censos (INEC)

En Ecuador el uso de la tecnología permite identificar la importancia del manejo de información para la comunicación.

Figura 2. Frecuencia del uso de Internet a nivel nacional

Fuente: Encuesta Nacional de Empleo Desempleo y Subempleo – ENEMDU (2015).

Elaboración: Instituto Nacional de Estadística y censos (INEC)

La Figura 2, muestra el cambio producido a nivel Nacional sobre el uso de internet, donde podemos apreciar que el 65,6% de las personas lo hacen por lo menos una vez al día, seguidos de los que por lo menos lo utilizan una vez a la semana con un 30,4%.

2.2 Brecha Digital

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

De acuerdo a las condiciones de vida y con los objetivos más relevantes se establecen enfoques que enriquecen el concepto determinando (Stillo, 2012):

- **Brecha digital de acceso:** generada por la inestabilidad económica, social o geográfica, es un enfoque que prioriza la infraestructura; es decir, la posibilidad o dificultad para disponer de computadoras conectadas.
- **Brecha digital de uso:** se refiere al desconocimiento de las tecnologías, pero con acceso. Se centra en el uso de las tecnologías y da referencia a la alfabetización digital.
- **Brecha digital de calidad de uso:** es para los usuarios que usan tecnologías de la información, pero con escaso acceso y sin mayor rendimiento.

  Esto se debe al uso de recursos disponibles en la red ya que se integran nuevos modos de trabajo, educación, negocios y entretenimiento, los mismos que pueden ser aprovechados con ciertos lenguajes o procesos de aprendizaje autónomos.

2.3 Consumo Digital

De acuerdo a un estudio realizado sobre el consumo de noticias (Casero-Ripollés, 2012) el consumo digital incluye un proceso de transformaciones debido al avance de la digitalización, concluyendo que las redes sociales son herramientas informativas donde se detecta un interés elevado de los jóvenes hacia las noticias y términos cívicos.

El consumo digital se presenta en;

- **Ambiente social.**

Según la Red Internacional (2016) las personas entre 18 y 24 años, utilizan redes sociales como su principal fuente de noticias con el 28% y el otro 24% televisión. Señalando que este consumo digital es frecuente con la plataforma Facebook con el 44 %, seguida de YouTube 19 %, Twitter 10 % y WhatsApp 8 %.

- **Ambiente político.**

Según El Mundo (2016), al definir públicos e identificar cuáles son los mensajes que interesan, la forma de localizarlos es por sus perfiles de Facebook .

Considerando el impulso digital y sincronizando mensajes adecuados a personas adecuadas en el momento adecuado, se pretende intercambiar información con los internautas por ser auténticos y compartir opiniones.

Según El Tiempo (2015) las redes sociales se utilizan de acuerdo a la situación que estén viviendo.

- Si están en un entorno muy concurrido, donde la diversión es la prioridad, suelen compartir sus momentos en WhatsApp, Facebook, Instagram, YouTube y Facebook Messenger.
- Si, por el contrario, desean compartir eventos más personales, en lugares no tan concurridos, acuden a Snapchat, Tinder y Vine.

“Las Nuevas Tecnologías de la Información y Comunicación (NTIC) aplicadas a la participación política admiten un mecanismo ciudadano con crecimiento sostenido”.(Resina, 2010)

- **Ambiente técnico.**

Las redes sociales se han convertido en espacios de instancias con el acceso a ideas y labores donde la interdependencia termina por generar influencia.

Para mayor circulación de información y de acuerdo al uso recíproco de tecnologías de información, se crean una red de intercambio de experiencias con apoyo en recursos electrónicos que permiten la acción clara y directa al conocimiento. (Lara & Duart, 2005).

2.4 Marketing Digital

Es la comunicación digital que permite formar en los medios sociales publicidad, generando un cambio en la relación social y proporcionando la comprensión y transmisión de información.

Permite al usuario que se comunique dinámicamente, pasando de ser un receptor a un coeditor de información, capturando así el interés personal y empresarial por brindar servicios de forma personificada y de acuerdo a las necesidades de cada usuario.

2.5 Redes Sociales

Son el medio de entretenimiento más accesible de la población en general, por ser una herramienta de comunicación rápida, con múltiples opiniones y varias fuentes de información, que pueden generar un gran valor y cumplir una variedad de exigencias que solucionan las necesidades de la sociedad actual.

Las redes sociales forman una estructura compuesta de personas conectadas a la infraestructura tecnológica, por una o varias relaciones de amistad, laboral, entretenimiento, entre otro tipo de interés que permiten la inserción de noticias, publicidad e información de manera adecuada, proporcionando la comprensión y entendimiento.

Según González (2015) las redes sociales contribuyen en varios aspectos como el comportamiento y las formas de relacionarse.

Dentro de las redes sociales se deben considerar los aspectos determinados a continuación para identificar riesgos o beneficios:

- **Aspectos Físicos:** actúan directamente a la integridad física de un usuario.
- **Psicológicos:** inciden en el comportamiento.
- **Económico:** con efecto positivo o negativo en el usuario y
- **Social**: que apoya o dificulta el desarrollo del usuario dentro de la sociedad.

2.5.1 Facebook.

Es una herramienta de comunicación masiva por la forma de segmentar la información, es un sitio web gratuito de redes sociales.

Se creó en 2004 como un hobby de Mark Zuckerberg que en aquel momento era estudiante de Harvard y como un servicio para los estudiantes de su universidad; dentro del perfil existen:

- **Lista de Amigos:** donde el usuario puede agregar a personas que conozca y estén registradas, siempre que acepte su invitación; para mantener un intercambio de fotos y mensajes; cuenta también con herramientas de búsqueda y de sugerencia de amigos.

- **Grupos y Páginas:** trata de reunir personas con intereses en común, donde se pueden añadir fotos, videos y mensajes. Las paginas se crean con fines específicos y los grupos están enfocados a marcas o personajes específicos con la prohibición de temáticas discriminatorias, con la opción de denunciar y reportar.

- **Muro:** existe en cada perfil y permite que los amigos escriban mensajes o intercambien información con imágenes o logotipos para que el usuario los vea.

- **Regalos:** son pequeños íconos con un mensaje que aparecen en la pared o se pueden enviar de forma privada que no es visible para todos los usuarios.

- **Juegos:** son aplicaciones relacionadas a juegos de rol, con pruebas de habilidades (digitación, memoria). Entre los más célebres se encuentran los juegos de Playfish, de Zynga Games como FarmVille10 y los juegos de Digital Chocolate como Tower Bloxx. (Mezrich & Vilà Vernis, 2010)

2.5.2 YouTube.

Es una herramienta que permite a sus usuarios descubrir, visualizar y compartir videos. Desde su lanzamiento, en mayo de 2005, ofrece un foro para que los usuarios se conecten, se informen e inspiren a otras personas en todo el mundo.

Es una de las empresas de Google, que cuenta con un reproductor online que permite la visualización de un archivo sin necesidad de descargarlo con la reproducción instantánea.

2.6 Trabajos Relacionados

- En Educación

Dadas varias investigaciones en países europeos, se determina que las redes sociales son de uso masivo, pero tienen un aprovechamiento limitado por la factibilidad y solución de problemas sin la guía de un docente.

Existen varios estudios que reflejan el efecto que producen las redes sociales, tal es el caso de una empresa de Hong Kong quien levantó una encuesta de 190 directivos, confirmando que los datos compartidos contribuyen a compartir conocimiento según los objetivos planteados.

- En la Sociedad

La información compartida permite el logro de metas de acuerdo a objetivos similares, generando éxito en la toma de decisiones organizacionales.

Las redes sociales generan influencia con una inscripción voluntaria, manteniendo la libertad de utilizar los medios tecnológicos para intercambios e integración de conocimiento, estrechando lazos y generando cambios.

Un estudio realizado para la gestión de reputación y medios sociales determina, que los adultos jóvenes de entre 18 a 29 años de edad son gestores en las redes sociales por la limitación de lo que comparten en línea a diferencia de los adultos mayores que no manejan permisos y hacen que su red social sea más accesible. (Madden & Smith, 2010)

(Bernete Francisco, 2010) menciona sobre la sociabilización de los jóvenes con el uso de las TIC, describiendo las consecuencias del uso de redes sociales, como:

- *“Lo público, en internet, se parece mucho a una extensión de lo privado en el sentido de que las comunidades que se forman son tantas y, a la vez, tan personales (basadas en intereses, afinidades, gustos propios) que tienen poco que ver con las estructuras sociales que coartan la libertad del individuo, refuerzan las relaciones de unos con otros; la condición de adultos, con los derechos y deberes de un ciudadano, sus miembros adquieren la satisfacción de sentirse*

integrados, pero serlo en uno de los miles de micro grupos sociales dice más bien poco o nada respecto de su integración en el conjunto social; y, en no pocas ocasiones, es una forma de segregación. La mayoría de estas "comunidades" son porosas y operan como vasos comunicantes entre ellas. Pero no sería prudente desconocer que también hay comunidades cuyos miembros perciben el conjunto social (instituciones y leyes) como algo ajeno a su vida y la vida de "los suyos", su cadena de amigos."

- En la política

Se usa como estrategia de comunicación en campañas políticas, como en los casos de Barack Obama y Donal Trump, quienes utilizaron la big data para generar variaciones en los mensajes y emitirlos de manera personalizada; la estrategia de Trump genero 175 mil versiones de un mismo mensaje con pequeños cambios en color, imagen y subtitulo, permitiendo así un cambio político de aceptación en la población.

López (2016) afirma que en las elecciones de España del 2015 y luego de analizar la actividad en Twitter *"Los resultados muestran a los candidatos de partidos emergentes, que tienden a enviar mensajes para movilizar a sus seguidores y para hacer anuncios genéricos que pronostican su futura victoria y la llegada de un cambio político, mientras que los líderes de los principales partidos políticos tienden a publicar más mensajes con propuestas programáticas específicas".* El manejo de datos políticos online, forman una serie de tendencias que pueden generar cambios en la sociedad. (Resina, 2010).

Según Gamir Ríos (2016) la presencia de candidatos y cargos públicos, forjan gran cantidad de información en redes sociales, donde se moviliza a los seguidores para la determinación de objetivos y toma de decisiones; la iteración en medios sociales permite establecer que "una imagen más que el contenido" por el impacto que generan en los usuarios al convertirse en el principal medio de comunicación. (Ruiz del Olmo & Bustos Díaz, 2016)

- En la Comunicación

Han generado un cambio, manejando perfiles responsables con destrezas y habilidades tecnológicas aplicadas a la población, con datos del 20% informativos y 80% de entretenimiento. Tal es el caso de España en Facebook que permitió adecuar al receptor como participante, dando protagonismo e importancia a la creación de discursos, facilitando la ejecución de planes estratégicos con la obtención del nuevo aprendizaje. (Barroso, Muñoz de la Luna, & Navarro, 2013)

2.7 Minería de Datos

Es una etapa del proceso de descubrimiento de conocimiento y permite tratar los datos para la obtención de patrones, perfiles, tendencias o modelos.

A fin de contribuir en la toma de decisiones se busca un uso compartido de información con la mejor comprensión del concepto. La minería de datos se presenta como una tecnología de apoyo para explorar, analizar, comprender y aplicar el conocimiento usando grandes cantidades de datos.

Los datos particulares permiten a la estadística cuantificar la incertidumbre de los patrones resultantes. (Asencios & Asencios, 2014)

Según Valencia (2012) se la usa como herramienta para la toma de decisiones estratégicas; considera a la información como un activo fundamental para las organizaciones, que al ser tratada con la minería de datos proporciona resultados y argumentos para la toma de decisiones.

(Microsoft, 2016) señala que la minería de datos o data mining es el proceso para detectar información en acción de grandes conjuntos de datos. Se apoya y utiliza el análisis matemático para la deducción de patrones existentes, los cuales no pueden ser detectados con la exploración tradicional por las complejas relaciones que presenta.

La minería de datos es una herramienta de extracción de información previamente desconocida y potencialmente útil, automatiza el proceso de búsqueda de relaciones y patrones en los datos; proporciona resultados válidos, novedosos y comprensibles.

- **Tipos de modelo.**

Se aplican luego del análisis de datos y se agrupan en:

- **Modelos Predictivos:** estiman valores desconocidos o futuros de variables de interés.
- **Modelos Descriptivos:** explora e identifican patrones de propiedades de datos que explican y resumen el comportamiento de determinada actividad.

Las técnicas aplicables a los modelos descritos se agrupan en:

2.7.1 Predictivas o supervisadas.

Hernández (2004) afirma: “Se trata de problemas y tareas en las que hay que predecir uno o más variables para uno o más problemas” (p.139). Estas tareas están orientadas

a describir un conjunto de datos que permiten obtener pronósticos de comportamientos futuros a partir de los datos recopilados, de ahí que se apliquen frecuentemente, se desglosan en Técnicas de predicción y asociación.

Seguidamente algunas de las técnicas más representativas:

- **Arboles de decisión:** son estructuras en forma de árbol que representan un conjunto de decisiones para la clasificación de un conjunto de datos.

  Un árbol de decisión es una estructura jerárquica formada por un conjunto de nodos, en donde cada nodo hace referencia a una regla o condición, que puede tener valores de verdadero o falso. De tal manera que la decisión final se pueda determinar siguiendo las condiciones desde el nodo raíz del árbol hasta los nodos hijos se cumple lo esperado.

  - **Algoritmo J48:** También conocido como algoritmo C45. Permite la predicción y clasificación basada en la teoría de la clasificación de datos, permite trabajar con valores continuos para atributos, separando los posibles resultados y dos ramas, para así escoger un rango de medida apropiado.
  - **Algoritmo REP Tree:** clasifica valores para atributos numéricos, permite construir un árbol de decisión usando información de varianza y se poda, usando como criterio de reducción de error. (Hernandez& Abilowo, 2016)

- **Métodos Bayesianos:** se caracterizan por el uso de distribuciones de probabilidad como una inferencia estadística en la que las evidencias se usan para cuantificar o actualizar una hipótesis con los datos que se quiera moldear.

Según Hernández (2004), es uno de los más usados en problemas de inteligencia artificial, aprendizaje automático y minería de datos para la inferencia de los datos.

El algoritmo de Bayes, permite estimar la probabilidad de pertenencia y es aconsejable utilizar grandes cantidades de datos para una predicción más correcta, ya que al tener poca información el modelo no puede ser adecuado. Los métodos bayesianos utilizan la tarea de clasificación y de acuerdo a los patrones de comportamiento existen algunos algoritmos que lo utilizan como:

  - **Algoritmo Naive Bayes:** se basa en un modelo de probabilidades que integra suposiciones de independencia y no tiene efecto sobre la realidad.

- **Redes Bayesianas (RBs):** “Es un formalismo que ha demostrado su potencialidad como modelo de representación del conocimiento con incertidumbre… Es una herramienta muy atractiva en su uso como representación del conocimiento, aspecto muy importante de la minería de datos” (Hernández, 2004) p. 263.

2.7.2 No supervisadas o descriptivas

- **Clustering (Agrupamiento)**

Es la que concentra datos dentro de un número de clases, partiendo de criterios de distancia o similitud, de manera que sean similares entre sí y distintas de las otras clases. Su utilización ha proporcionado significativos resultados en lo que respecta a la clasificación de patrones; estos pueden ser:

- **Redes neuronales:** basada en la forma en la que funciona el sistema nervioso de los animales, permite interconectar las neuronas en una red que presta colaboración para la producción de estímulos.

- **Algoritmo Kmeans:** es uno de los más conocidos en donde se divide la data en K grupos, su idea principal es la definición del k (centroide), toma los objetos y se ubica en su centroide más cercano, luego recalcula el centroide de cada grupo y vuelve a distribuirlos de acuerdo al centroide más cercano.

- **Cobweb:** es un algoritmo jerárquico que utiliza el aprendizaje incremental, realiza agrupaciones de instancia a instancia; genera un árbol donde las hojas representan segmentos y el nodo raíz contiene el conjunto de datos de entrada.

  Las instancias se adicionan y se actualizan al encontrar el mejor lugar, así reconstruyen el árbol considerando la utilidad de la categoría que mide la calidad de una partición de instancias en un segmento. Este algoritmo considera dos parámetros:
  - **Acuity:** utilidad de la categoría que utiliza la estimación de la media y desviación estándar de los atributos.
  - **Cut-off:** es para evitar el crecimiento del número de segmentos.

- **Redes de Kohonen:** llamado también mapa auto organizado o SOM (Self-Organizing Maps), cuyo proceso es similar al del cerebro, ya que determina

rasgos comunes, regularidades en datos entrantes para incorporarlos y organizarlos a su estructura de conexiones en función de los datos procedentes del exterior.

Al formarse la red se activa un patrón de entrada, donde sólo una de las neuronas de salida (o un grupo de vecinas) se activa, por lo que las neuronas compiten, dando a una como neurona vencedora y anula al resto, por valores mínimos.

- **Algoritmos genéticos y evolutivos:** son un método de búsqueda que utiliza la optimización de procesos imitando la evolución de la teoría biológica de Charles Darwin. (Peña & Parra, 2016)

- **Máquinas de vectores de soporte:** o SVM (Support vector machines) se basan en el aprendizaje estadístico y usan datos dimensionales, con este se pretende encontrar hiperplanos óptimos que separan un conjunto de datos en clases. (Ramírez, 2007).

- **Asociación**

Permiten establecer las posibles relaciones o correlaciones entre distintas acciones o sucesos aparentemente independientes; pudiendo reconocer como la ocurrencia de un suceso o acción puede inducir o generar la aparición de otros.

(Hernández et al., 2004) afirma que las reglas de asociación son una manera de expresar patrones de datos de una base de datos. Estos patrones pueden servir para conocer el comportamiento general de un problema generado por la base de datos, considerándolo para la toma de decisiones. Dentro de los algoritmos más usados está el A priori.

Al trabajar con las reglas de asociación se utiliza dos medidas para conocer la calidad de la regla, cobertura (support) y confianza (confidence).

La regla predice correctamente el número de instancias y estas son la cobertura, mientras que la confianza (conocida como precisión) es aquella que mide el porcentaje de veces que se cumple la regla y cuando se puede aplicar. (Hernández et al., 2004).

- **Algoritmo A priori:** “Se basa en la búsqueda de los conjuntos de ítems con determinada cobertura. Para ello se construyen los conjuntos formados por un solo ítem que supera la cobertura mínima. Este conjunto de datos se utiliza para construir el conjunto de conjuntos de dos ítems y así sucesivamente hasta que

llegue a un tamaño en el cual no existan conjuntos de ítems con la cobertura requerida". (Hernández et al., 2004) p.240.

Principalmente se basan en el conocimiento previo o "a priori" de los conjuntos frecuentes, se utiliza para reducir el espacio de búsqueda y aumentar su eficiencia.

- **Regresión Logística:** modelo de regresión lineal simple, es quien predice la probabilidad que un resultado pueda tener dos valores (dicotómica o binaria) en función de las variables independientes, las cuales pueden ser cualitativas o cuantitativas, adoptando posibles valores: 1 y 10, positivo y negativo, éxito y fracaso, etc.

Tabla 1 Algunas Técnicas de Minería de Datos

Nombre	**PREDICTIVO**		**DESCRIPTIVO**		
	Clasificación	**Regresión**	**Agrupamiento**	**Asociación**	**Correlaciones/ Factorizaciones**
Arboles de decisión ID3, C5.0	✓				
Arboles de decisión CART	✓	✓			
Otros Arboles de decisión	✓	✓	✓	✓	
Redes Neuronales	✓	✓	✓		
Redes de Kohonen			✓		
Regresión lineal Y Logarítmica		✓			
Regresión Logística	✓			✓	
A priori				✓	

Kmeans			✓		
NaiveBayes	✓				
Vecinos más próximos	✓	✓	✓		
Análisis factorial y de componentes principales					✓
Twostep, Cobweb			✓		
Algoritmos Genéticos y evolutivos	✓	✓	✓	✓	✓
Máquinas de vectores de soporte	✓	✓	✓		
CN2 rules	✓			✓	
Análisis discriminante multivariante	✓				

Fuente: (Hernández et al., 2004)
Elaboración: propia

2.7.3 Proceso de descubrimiento de conocimiento en Bases de Datos

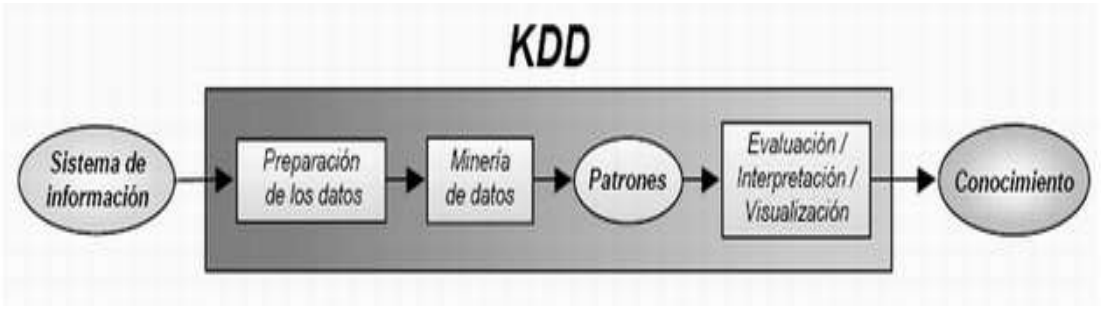


Figura 3. Proceso de Knowledge Discovery in databasecs
Fuente: (Hernández, Cèsar, & Ramírez, 2007)
Elaboración: (Hernández et al., 2007)

Como se muestra en la Figura 3, de la obtención de información y el proceso para el descubrimiento de un nuevo conocimiento y con las fases descritas en el proceso KDD, es posible extraer un conocimiento valido y útil. Las fases KDD se clasifican en:

2.7.3.1 ***Fase Integración y Recopilación***

En esta fase se integra la información recopilada, que a su vez esta tratada y unificada, dando como resultado un almacenamiento de datos que será útil para agregar y cruzar la información.

2.7.3.2 ***Fase de Selección, limpieza y transformación***

En esta fase se mejora la calidad y se elige datos y variables para la resolución del problema planteado.

Los problemas en esta fase se hacen presentes cuando:

- Los valores no se adaptan al comportamiento general de los datos
- Existen valores inconsistentes, irrelevantes y datos perdidos

De acuerdo a los modelos y herramientas se tratan estos problemas y se pueden solucionar de la manera requerida.

2.7.3.3 ***Fase de Minería de Datos***

En esta fase se añade un nuevo conocimiento en base a la información obtenida hasta el momento, para construir un modelo con patrones y relaciones de datos para entender los datos y explicar sucesos ocurridos, dando así la proyección para predecir situaciones y hechos relevantes.

Para la construcción del modelo es necesario considerar:

1. Determinar el tipo de tarea de la minería
2. Adoptar el modelo adecuado y correctamente el algoritmo de minería en base a la tarea y obtener el tipo de modelo que se necesita

2.7.3.4 ***Fase de Evaluación e Interpretación***

En esta fase se mide la calidad de los patrones descubiertos por los algoritmos de minería de datos, se deben tener en cuenta tres cualidades específicas:

- Patrones precisos, Comprensibles e Interesantes

2.7.4 Herramientas para minería de datos

Permiten convertir los datos en conocimiento dando así una ventaja de información. Existen algunas herramientas de código abierto entre las cuales se puede mencionar:

2.7.4.1 *Weka1 (Waikato Environment for Knowledge Analysis)*

Es una herramienta visual bajo la licencia general publica GNU desarrollada en Java, creada por investigadores de la Universidad de Waikato de Nueva Zelanda, contiene un conjunto de algoritmos de aprendizaje automático para tareas de minería de datos. Contiene interfaz gráfica para facilitar su uso. (Weka, 2011)

Permite un proceso previo, clustering o generación de grupos de datos, clasificación, regresiones, visualización y selección de propiedades. Su técnica se basa en la hipótesis que los datos se encuentran disponibles en un solo archivo o relacionados, donde cada dato está ubicado de acuerdo a su atributo, proporciona acceso a base de datos SQL, gracias a la conexión JDBC[2], que mediante una consulta permite procesar el resultado.

Características:

- Está constituido por paquetes de código libre como técnicas de pre-procesamiento, agrupamiento, clasificación y visualización.
- Los datos e pueden añadir con archivos en formato csv, arff, c4.5.
- No guarda parámetros de escala para aplicar a datos futuros
- Se integra en otros paquetes Java.

2.7.4.2 R Project[3]

Es una herramienta de software libre que maneja con un entorno de gráficos estadísticos. Es la unión de un conjunto de servicios para el manejo de datos, cálculos y representación gráfica, resulta muy útil al trabajar en diferentes plataformas Unix, Windows y Mac OS. Ofrece un ambiente dinámico y con una amplia variedad de técnicas y gráficas, incluyendo pruebas clásicas, agrupación, modelos lineales y no lineales, entre otros.

Trabaja con comandos de programación estadística y utiliza un lenguaje orientado a objetos.

---

[1] http://www.cs.waikato.ac.nz/ml/weka/
[2] Java Database Conectivity
[3] R Project: https//www/r-project.org/

CAPITULO III

3. METODOLOGIA

En el presente capítulo se detalla el método para la obtención de resultados del Trabajo de Titulación de acuerdo a las fases y técnicas a utilizar de la metodología KDD descritas en la sección 2.7.3 del capítulo II.

3.1 Muestreo de Datos

La información dentro de las redes sociales Facebook permite determinar la preferencia de contenido e identificar un patrón de uso de los habitantes de la provincia de Loja.

Para la muestra que es de tipo probabilístico, se utiliza la Ecuación1. Por utilizar datos de una población finita.

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

Ecuación 1. Fórmula para la muestra de datos
Fuente: (Herrera Castellanos, 2016)
Elaboración: propia

Donde:

n = tamaño de muestra.

N = valor de población. 214.855 (INEC, 2010)

Z= valor critico de coeficiente de confianza 95% = 0.95.

d = Margen de error aceptado 5% = 0.05.

p = proporción de ocurrencia de evento = 0.25.

q = proporción de no ocurrencia de evento= 0.25.

Al usar la Ecuación 1 y reemplazando valores se obtiene como resultado una muestra inicial de n = 139 (perfiles de usuarios).

Para el desarrollo del TT se utilizarán 500 (perfiles de usuarios) que es un valor mayor a la muestra solicitada con las ultimas 20 publicaciones de su perfil.

3.2 Proceso de Investigación

Para el desarrollo del Trabajo de Titulación se aplican las fases del KDD y de acuerdo a la información recolectada de los perfiles de usuarios de Facebook en la provincia de Loja.

3.2.1 Fase Integración y Recopilación.

Para iniciar con el Trabajo de Titulación, la captura de perfiles se realizó mediante un método sistemático aleatorio, que consiste en la selección aleatoria de un perfil de la red social y la elección del décimo usuario del perfil agregado, considerando la edad y localidad.

Se empieza a construir una base de datos con la captura del perfil, su género y su edad como se muestra en la Figura 4.

Perfil de Facebook	Edad	Genero
https://www.facebook.com/profile.php?	15	F
https://www.facebook.com/profile.php?	16	F
https://www.facebook.com/profile.php?	17	M
https://www.facebook.com/profile.php?	18	F
https://www.facebook.com/profile.php?	19	M
https://www.facebook.com/profile.php?	20	M
https://www.facebook.com/profile.php?	21	F
https://www.facebook.com/profile.php?	24	M
https://www.facebook.com/profile.php?	25	M
https://www.facebook.com/profile.php?	27	M
https://www.facebook.com/profile.php?	28	F
https://www.facebook.com/profile.php?	30	F
https://www.facebook.com/profile.php?	31	M
https://www.facebook.com/profile.php?	37	M
https://www.facebook.com/profile.php?	38	F
https://www.facebook.com/profile.php?	39	F
https://www.facebook.com/profile.php?	40	F
https://www.facebook.com/profile.php?	41	M
https://www.facebook.com/profile.php?	42	M
https://www.facebook.com/profile.php?	47	M
https://www.facebook.com/profile.php?	50	M
https://www.facebook.com/profile.php?	51	F
https://www.facebook.com/profile.php?	52	F
https://www.facebook.com/profile.php?	62	F
https://www.facebook.com/profile.php?	66	M
https://www.facebook.com/profile.php?	70	M
https://www.facebook.com/profile.php?	71	F

15-18
19-24
25-30
30-40
40-50
50 a mas

Figura 4.Muestra de perfiles categorizados por edad.
Fuente: La autora
Elaboración: propia

Una vez realizada esta primera recolección de datos, se accede a cada uno de los perfiles añadidos y se considera:

- Información de las últimas 20 publicaciones y
- El contenido en cada perfil para la determinación del tipo de contenido.

En primera instancia y de acuerdo a la información recolectada se determinas el tipo de contenido descritos en la Tabla 2.

Tabla 2 Categorías de Perfiles de Usuarios

CATEGORIAS	
Naturaleza	ubicación geográfic-a
Noticias	Animación
Decoración	Horóscopo
Arte	Belleza
relación sentimental	Moda
Animales	Música
Política	Humor
Religión	personal selfie
Familia	Cultura
Amistad	Tecnología
Reflexión	Educación
Comida	Estado
Manualidad	Deportes
Solidaridad	Información
Felicitaciones	Salud
Entretenimiento	

Fuente: La autora
Elaboración: propia

Para la recolección de información se manejan datos con las variables descritas en la Tabla 3.

Tabla 3. Campos Base de Datos

Campo	Variable	Tipo de Dato	Descripción	Valores
Nombre	Nominal	Cadena	Perfil de Usuario	Ninguno
Edad	Real	Numérico	Edad del Usuario	Grupo 1=15-18 Grupo 2=19-24 Grupo 3=25-30 Grupo 4=30-40 Grupo 5=40-50 Grupo 6=50 a más.
Genero	Dicotómica	Numérico	Género del Usuario	Femenino Masculino
Tipo de contenido	Nominal	Numérico	Nombre para post de las publicaciones en el Perfil	Ninguno

Fuente: La autora
Elaboración: propia.

3.2.2 Fase de Selección, limpieza y transformación.

Para esta fase se realiza un control de los datos obtenidos en la fase anterior y se selecciona los perfiles de usuario determinados mediante un método sistemático aleatorio, que consiste en la selección aleatoria de un perfil de la red social y la elección del décimo usuario del perfil agregado y con 6 grupos de edad.

Se busca un manejo de datos de forma eficiente para lo cual se realiza la limpieza y la verificación de datos faltantes y se usan las 8 categorías más significativas determinadas en la Tabla 4.

Tabla 4. Categorías para minería de datos

CATEGORIAS
personal selfies

reflexión
Familia
Amistad
Humor
Música
Estado
relación sentimental

Fuente: La autora
Elaboración: propia

3.2.3 Fase de minería de datos.

En esta fase se extrae el conocimiento en base a los datos obtenidos en la fase anterior, esta información permite construir un modelo formado por patrones y relaciones de los datos obtenidos.

Para esto se debe definir el modelo y el algoritmo de minería de datos a utilizar, esta fase se la llevará a cabo usando el lenguaje de programación de R Project definido en el capítulo II, sección 2.7.4.2.

3.2.3.1 Identificación de modelo.

Para esto se considera los objetivos determinados en el capítulo I.

En el Trabajo de Titulación se utilizará un modelo descriptivo que permitirá utilizar las técnicas y herramientas adecuadas para la descripción del perfil de uso en base al contenido de los perfiles de redes sociales de los usuarios de la provincia de Loja.

3.2.3.2 Algoritmo de minería de datos.

La tarea seleccionada para el modelo definido anteriormente es:

- **Clusterización:** es la que permite obtener grupos a partir de características similares descritas en la base de datos.

En el Trabajo de Titulación se utilizará la técnica de clustering con el algoritmo Kmeans, mapas de Kohonen y arboles de decisión descritos en el capítulo II, sección 2.7.2.

Se implementa un análisis estadístico que permitirá descubrir características similares en base al contenido determinado por la organización de los datos.

- ✓ ***Mapas de Kohonen***

Como simula el funcionamiento del cerebro, consta de dos capas con neuronas de entradas y M de salida, las conexiones de salida se las recibe con pesos, cuando la red evoluciona sólo una neurona de salida se activa quedando como la neurona vencedora.

Para los datos de entrada, primero se debe tener un conjunto de datos entrantes(Merelo, J, 2004). A continuación, se detallan los pasos a seguir:

1) Primero se inicializan los pesos, con valores aleatorios pequeños, W_{ji}
2) La información de entrada se presenta en forma de vector, $E_k = (e_1 \dots\dots e_n)$, donde e son valores continuos.
3) Se realiza la determinación de la neurona vencedora de la capa de salida, en donde el vector de pesos W_j, será el más aproximado a la entrada E_k. Para esto se realiza el cálculo de las distancias entre los vectores mencionados. En la Ecuación 1, se muestra uno de las más usadas la distancia euclidiana.

$$d^2(W_{ij}, X) = \sum_{k=1}^{n} (W_{ijk} - X_k)^2$$

Ecuación 2. Fórmula para calcular la distancia euclidiana.
Fuente: (Flórez López & Fernández Fernández, 2008)
Elaboración: propia

4) Para determinar la neurona ganadora se considera a aquella cuya distancia es la menor de todas.
5) Cuando se tiene la neurona ganadora se realiza la actualización de pesos de esta neurona y de la de sus vecinas, para esta actualización se usa la distancia euclidiana(DE) como se muestra en la Ecuación 3:

$$W_{ij}(t+1) = W_{ij}t + \beta(t)(e_i{}^k - W_{j*i}(t))$$

Ecuación 3. Actualización de pesos, similar a distancia euclidiana
Fuente: (Flórez López & Fernández Fernández, 2008)
Elaboración: propia

, donde $\beta(t)$ es un parámetro llamado ritmo de aprendizaje, se puede expresar como $\beta(t)\frac{1}{t}$.

6) Al realizar todas las iteraciones, se termina el proceso, en el caso contrario se vuelve al paso 2. (Flórez López & Fernández Fernández, 2008)

✓ ***Kmeans***

De acuerdo a (Rui Xu, 2009) los pasos a seguir para la ejecución del algoritmo son:

1) Inicializar cada patrón K de forma randómica o en base a centroides preexistentes. Calcular la matriz prototipo de clúster M= $[m_1, \ldots., m_k]$.
2) Asignar cada sujeto al clúster más cercano.
3) Re calcular la matriz prototipo del clúster en base a la partición actual.
4) Repetir los pasos 2 y 3 hasta que no existan cambios en el centroide del clúster.

Para la ejecución del algoritmo K-means se escoge las categorías establecidas en la Fase 3.2.1.

Como se muestra en la Figura 6, para el método de k-means se utiliza el codo de jambu que depende de una selección inicial de particiones y hace uso de los algoritmos determinados, siendo el "Hartigan-Wong" el más adecuado para la data obtenida.

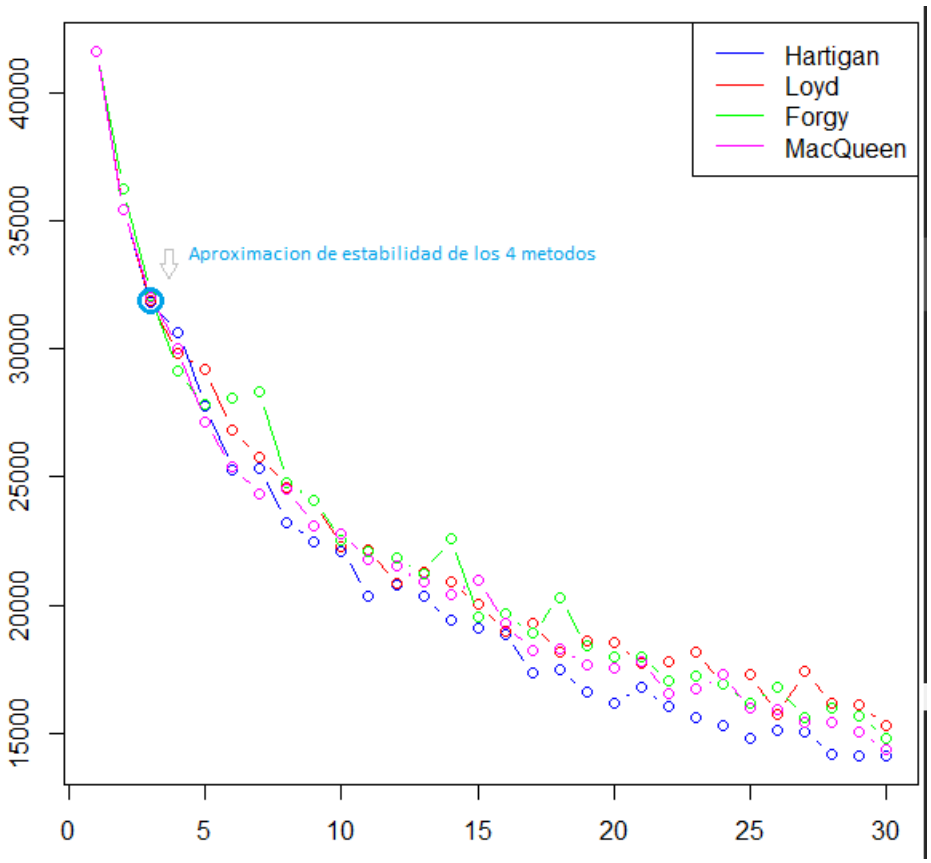


Figura 5. Algoritmo Hartigan (Codo de Jambu)
Fuente: La autora
Elaboración: propia

La Figura 5, describe la aproximación de la estabilidad de los algoritmos para la calibración y selección del modelo de forma que maximicen la inercia Inter-Clases y minimicen el error global.

Con el método Kmeans es posible ejecutar los algoritmos: Hartigan-Wong, Lloyd, Forgy y MacQueen. De acuerdo a los datos recolectados y para la elección del modelo con la mejor distribución de los datos, el mejor algoritmo a aplicar es el Hartigan por la mayor distancia entre número de clusters que presenta, porque mientras mayor es la distancia inter-clases, mejor es la clusterización.(Rodríguez, 2014)

3.2.4 Fase de Evaluación e Interpretación

Para esta fase se verifican y determinan los patrones obtenidos en la fase anterior, se utilizan algunas medidas y técnicas de minería de datos para conocer la obtención de resultados.

CAPITULO IV

4. RESULTADOS

En este capítulo se muestran los datos generales, la descripción de contenido con los resultados y la discusión luego de la ejecución de los algoritmos determinados.

4.1 Fase de integración y Recopilación

Datos Generales

La Figura 6, indica los valores de la variable genero de los datos recolectados.

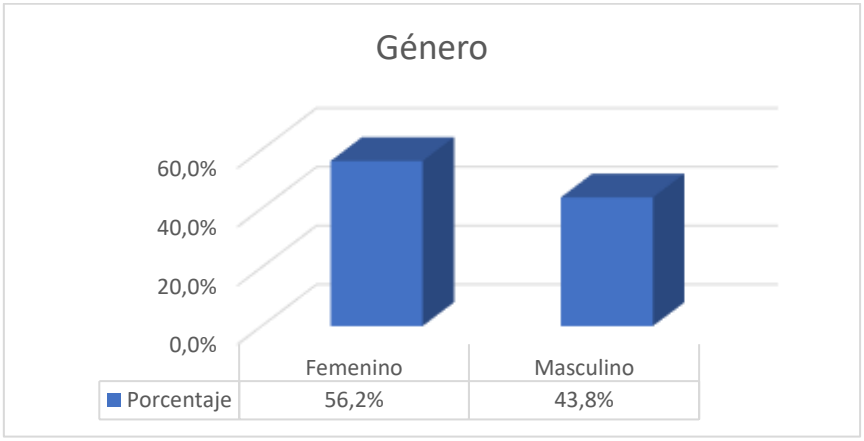


Figura 6. Genero Usuarios
Fuente: La autora
Elaboración: propia

Existe un amplio rango de edades presente en la muestra de los habitantes de la provincia de Loja como se muestra en la Figura 7.

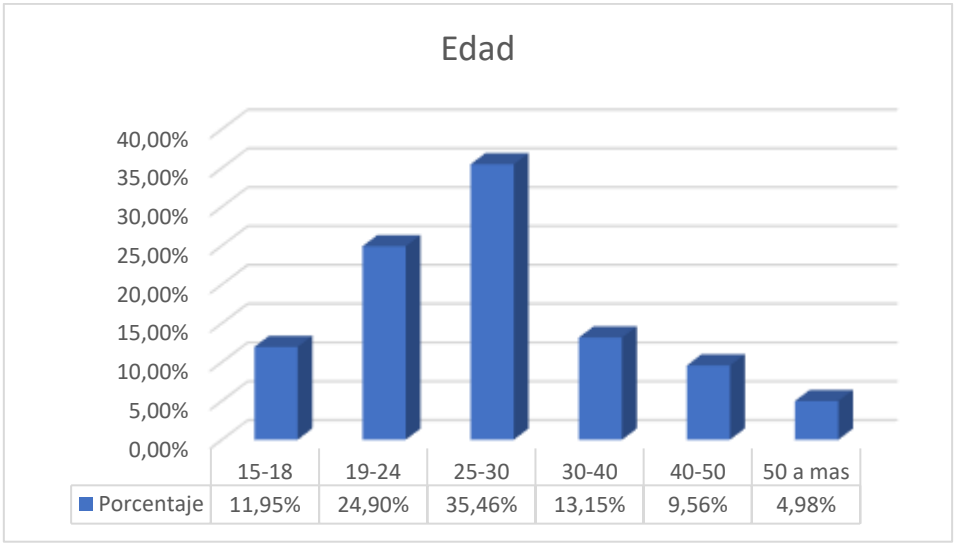


Figura 7. Edad de usuarios
Fuente: La autora
Elaboración: propia.

4.2 Fase de Selección, Limpieza y Transformación

Descripción de Contenido

El tipo de contenido es la clasificación de las publicaciones que cada usuario comparte dentro de su muro de red social y se determina en las siguientes:

- **personal selfies:** se refiere a las fotos personales, conocidas como selfies.
- **reflexión:** son imágenes o pensamientos que llevan a la reflexión de los hechos en los que se desenvuelven las personas en la sociedad.
- **familia:** es la información, mensajes, videos o fotos sobre la familia.
- **amistad:** son las publicaciones de amigos como videos, imágenes o pensamientos.
- **humor:** hace referencia a los posts de entretenimiento y risas con memes que representan la realidad con humor.
- **música:** información de cantantes, videos y conciertos.
- **estado:** son las publicaciones personales de ideas, estados de ánimo, emociones y hechos importantes.
- **relación sentimental:** se refiere a la información personal sobre sus parejas con imágenes videos o pensamientos.

La Figura 8, muestra el porcentaje de usuarios y el tipo de contenido, describiendo a los posts obtenidos en la captura de datos.

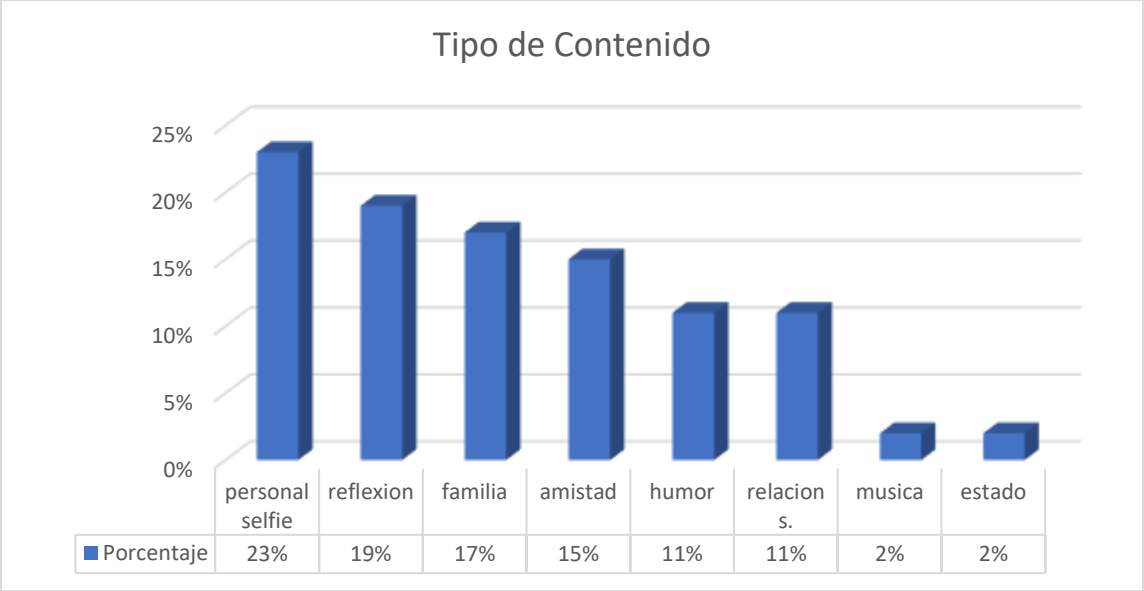


Figura 8. Tipo de Contenido
Fuente: La autora
Elaboración: propia

4.3 Fase de minería de datos

Clasificación con mapas de Kohonen

Con el tipo de contenido determinado se realiza la clasificación con mapas de Kohonen que tiene mayor énfasis en la visualización de los datos.

Para esta experimentación se utiliza la topología hexagonal de 5 columnas y 4 filas como se muestra en la Figura 9.

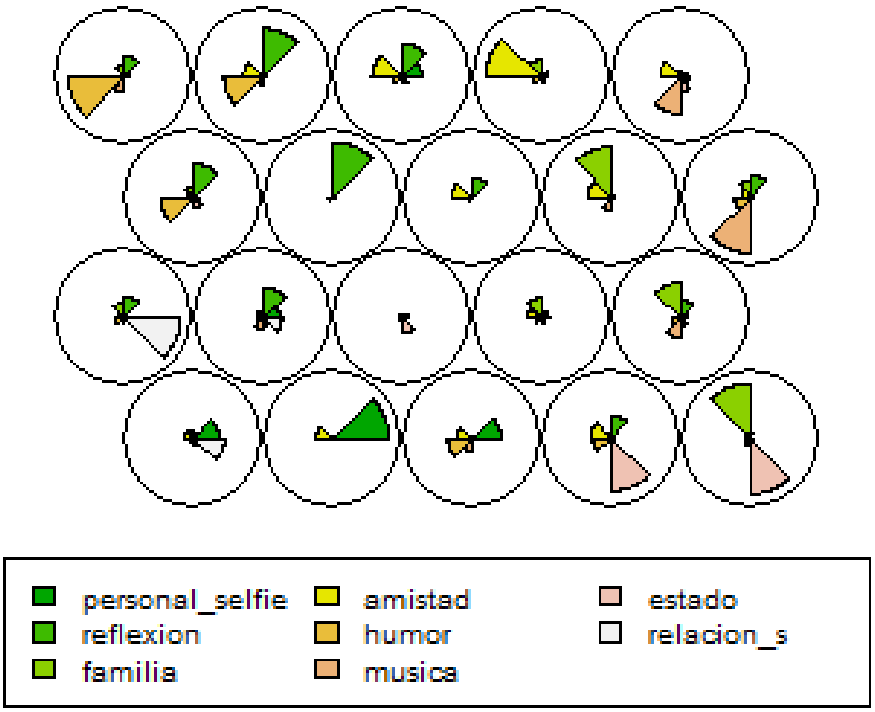


Figura 9. Mapas de Kohonen
Fuente: La autora
Elaboración: propia

En ella se muestran 20 nodos o neuronas, que surgen de realizar el modelo SOM (Self-Organizing Maps), que permite observar las agrupaciones dependiendo del nivel de interacción que posee cada categoría, es decir cada nodo o neurona muestra la utilización de categorías, en el nodo 3 la iteración es mínima por lo que las categorías son personal selfies, reflexión y amistad, mientras que en la neurona 20 las categorías son personal selfies y estado, pero es en la neurona 7 que resulta la categoría personal selfies como la categoría vencedora con la identificación de preferencia de contenido de los habitantes de la provincia de Loja.

El objetivo de esta clasificación, es utilizar valores similares y, por tanto, activar la misma neurona de salida.

Clasificación con algoritmo K-means

Para mejorar la fiabilidad del proyecto se realizaron varias clasificaciones, con el objetivo de utilizar la mejor clasificación se crearon clasificaciones para 2,3 y 4 grupos descritos en el Anexo II, III y IV.

Tabla 5. Centros finales de 4 clúster

TIPO DE CONTENIDO EN USUARIOS DE LA PROVINCIA DE LOJA	Clasificación por Grupos			
	1	**2**	**3**	**4**
personal selfies	1	1	2	9
reflexión	1	8	3	1
familia	4	2	1	2
amistad	2	1	1	3
humor	1	2	7	1
música	1	1	1	1
estado	1	1	1	1
relación sentimental	1			1

Fuente: La autora
Elaboración: propia

En la clasificación se utilizaron las variables de los centros finales descritas en la Tabla 7 y se puede determinar que la mejor clasificación para los usuarios de la muestra obtenida es de 4 grupos. Esta clasificación tiene algunas características entre ellas:

- Similitud en los datos recolectados
- Agrupación por tipo de contenido
- Existe una pequeña variación de datos dentro de un clúster con relación a otro clúster

La Figura 10, muestra la clasificación de los usuarios de la muestra obtenida de acuerdo a los 4 grupos determinados.

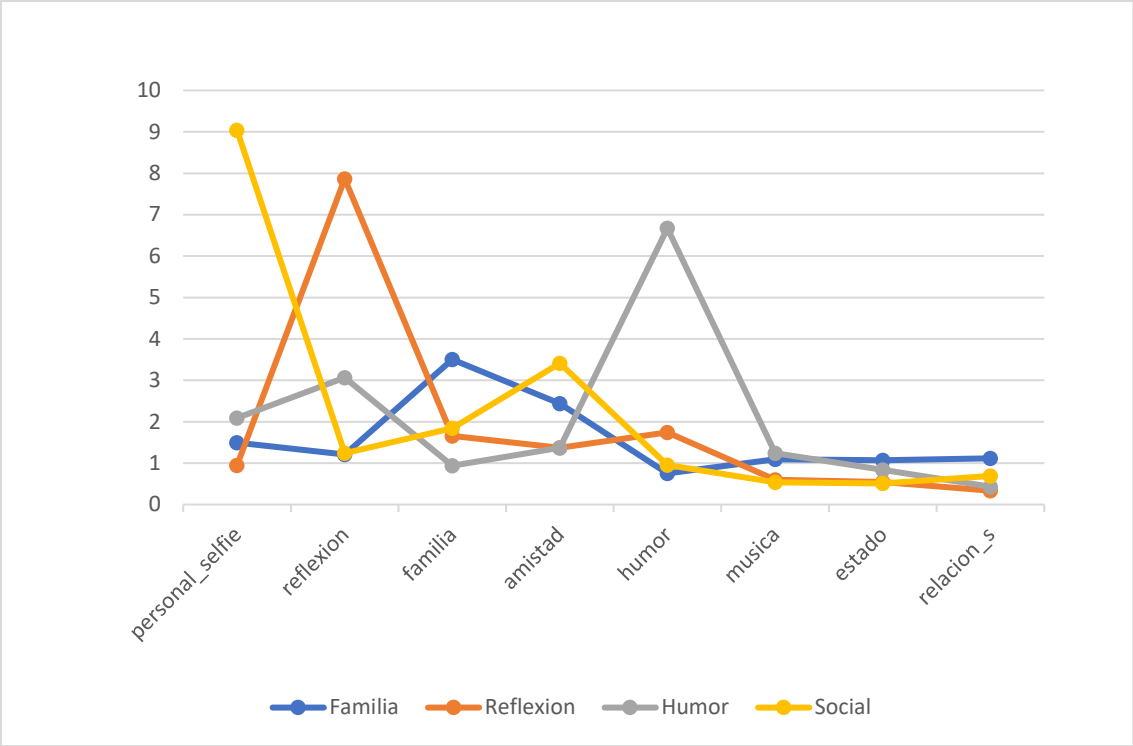


Figura 10. Clasificación en 4 grupos
Fuente: La autora
Elaboración: propia.

La distribución espacial que se puede observar en la Figura 11 de los 4 grupos, permite evidenciar que la distancia que se maneja para hacer que se pertenezca o no a cualquier grupo es mínima, sin embargo, cada usuario es asignado iterativamente al centroide más cercano del grupo, descartando diferencias importantes.

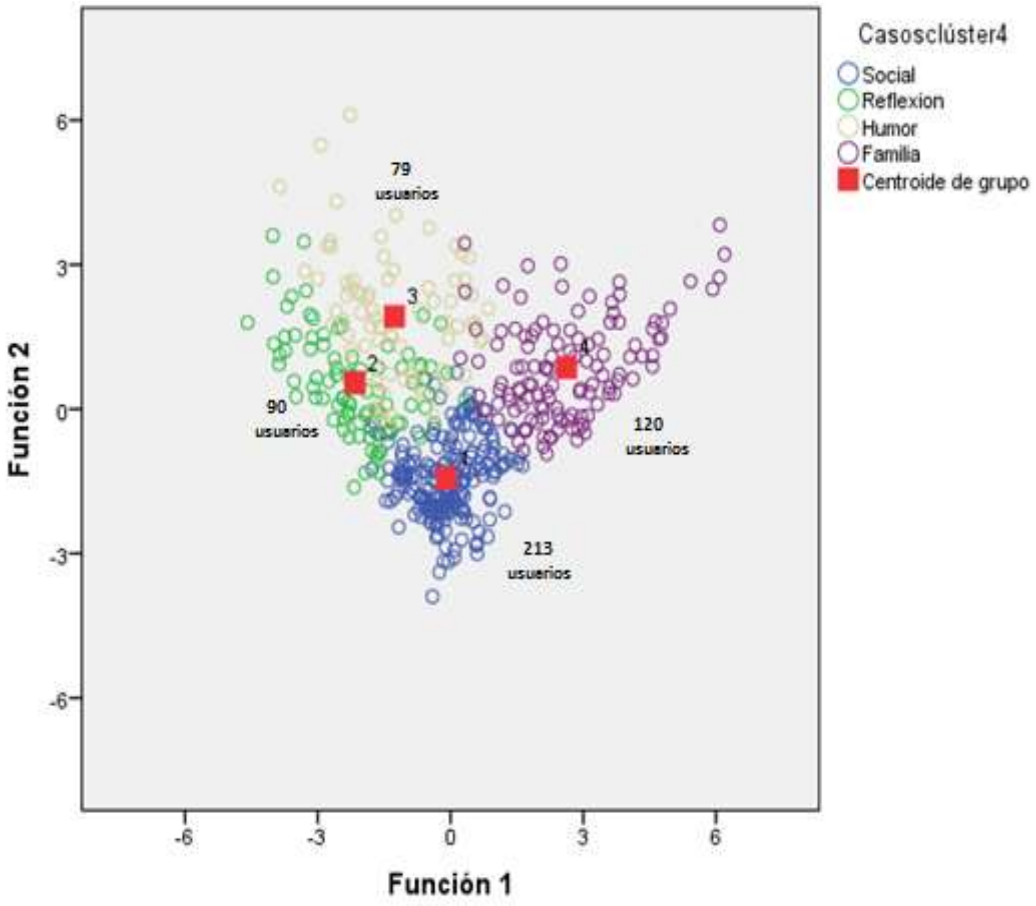


Figura 11. Clasificación en 4 grupos
Fuente: La autora
Elaboración propia.

Con la clasificación en 4 grupos es posible determinar y describir a los siguientes:

- **Grupo social:** contiene información social, personal y de relaciones afectivas de amistad, con imágenes, videos, comentarios y mensajes que estrechan lazos dentro de la sociedad, con información de hechos o sucesos importantes que brindan un direccionamiento de preferencia de contenido, incluyen:
    - Imágenes con fotos personales (selfies)
    - Momentos con amigos
    - Reuniones sociales (fiestas, celebraciones)
    - Eventos culturales
    - Eventos sociales
    - Deportes y noticias

- **Grupo reflexión:** comprende información que lleva a la meditación de situaciones, hechos y momentos personales que contribuyen a la generación de un nuevo conocimiento e incluyen:

- Imágenes, videos reflexivos
- Consejos
- pensamientos y
- reflexiones de vida

- **Grupo familia:** determinada por el contenido de relación directa con familiares que consta de:
  - Imágenes, videos familiares
  - Paseos
  - Celebraciones y
  - Momentos importantes con estrecha relación personal y familiar.

- **Grupo humor:** contempla información de entretenimiento con datos de juegos y bromas que se distingue con:
  - Imágenes denominadas memes (situaciones y hechos reales con humor) de deportes, política, educación, salud y religión.
  - Bromas, chistes y
  - Hechos sociales reflejados con ironía.

  En este grupo los usuarios ocupan el 16% de la población con esta preferencia de contenido.

Clasificación por preferencia de Contenido

También se aplica la técnica de árboles de decisión con el objetivo de determinar las opciones de preferencia de contenido de los habitantes de la provincia de Loja al mostrar las decisiones y resultados de la información analizada, con los 4 grupos de contenido ya establecidos.

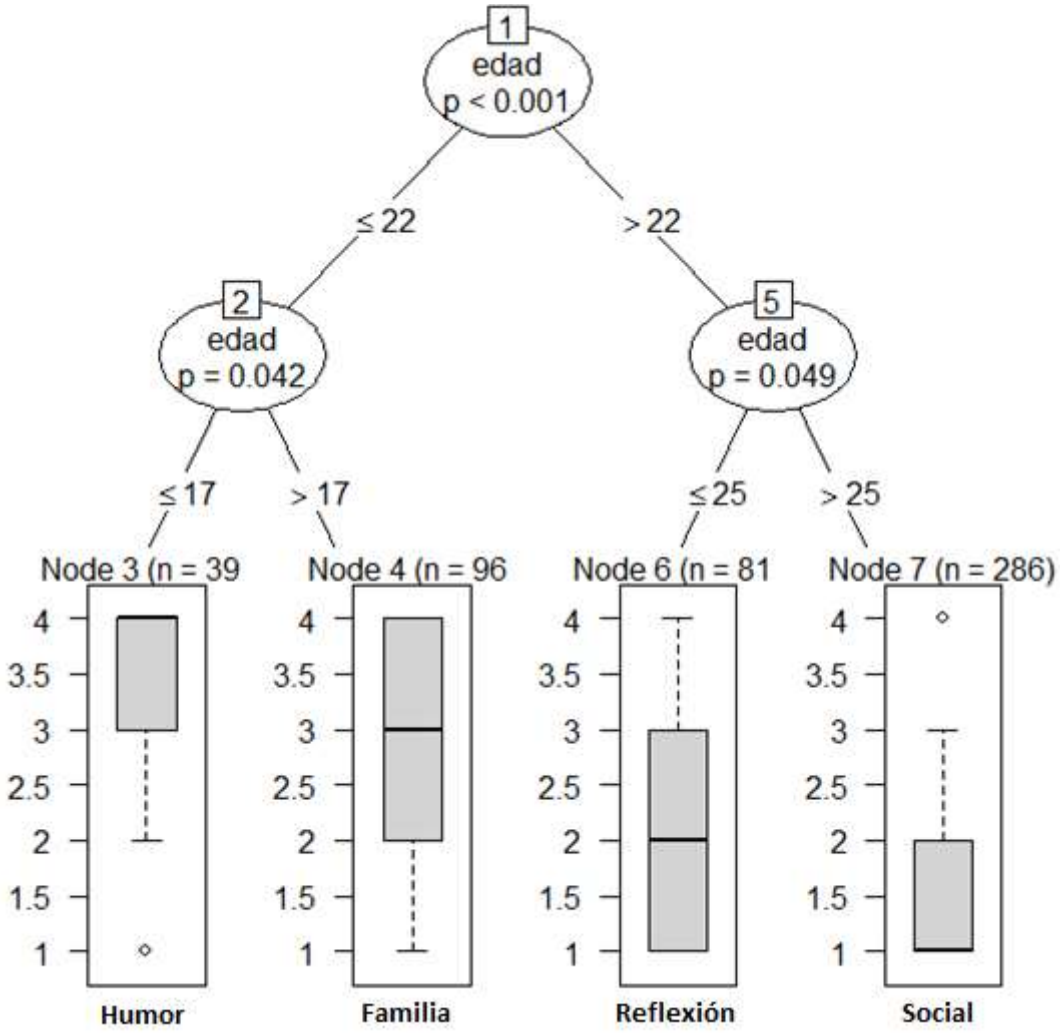


Figura 12. Árbol de Decisión
Fuente: La autora
Elaboración: propia

La Figura 12, identifica a los 4 grupos de contenido con la información de las publicaciones de la muestra recolectada, determinando las condiciones para la decisión de preferencia de contenido, de acuerdo a la edad de los usuarios, donde:

- Los usuarios menores o iguales a 22 años de edad identifican que
    - La elección o preferencia de contenido de los usuarios menores o iguales a 17 años es la del contenido del Grupo Humor con n=39.

- Para los usuarios mayores a 17 años de edad, la elección y preferencia de contenido es la del contenido del Grupo Familia con n=96.

- Los usuarios mayores a 22 años de edad identifican que las reglas de decisión se dan para:
  - Los usuarios menores o iguales a 25 años de edad eligen y prefieren el contenido del Grupo Reflexión con n=81.
  - Y para los usuarios mayores a 25 años de edad su elección de contenido es la del Grupo Social con n=286.

Al recorrer el árbol de decisión con los nodos que representan la clasificación de los usuarios por edad y los bordes con la preferencia de contenido, es posible determinar los perfiles de uso de las redes sociales en la población de la provincia de Loja.

Preferencia de Contenido por Edad

Con el cruce de variables y para el desarrollo de un nuevo conocimiento, se describe la preferencia de contenido por edad con la información descrita en la Tabla 8.

Tabla 6. Variables Edad- usuarios y grupos de contenido

	Edad					
Grupos de Contenido	**15-18**	**19-24**	**25-30**	**30-40**	**40-50**	**Más de 50**
SOCIAL	13%	30%	49%	65%	50%	56%
FAMILIA	60%	34%	13%	15%	15%	8%
REFLEXION	14%	13%	22%	14%	22%	28%
HUMOR	13%	23%	16%	6%	13%	8%
Total	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**

Fuente: La autora
Elaboración: propia

La clasificación en 4 grupos, permite identificar a los usuarios con la preferencia de contenido, considerando la edad de la muestra obtenida, es posible definir lo siguiente:

- El Grupo Social se describe con:
  - Los usuarios de **15 a 18** años de edad, que ocupan el **13%** de la muestra**,** donde las mujeres ocupan el **5%** de las publicaciones y los hombres el **8%** de los posts.
  - Los usuarios de **19 a 24** años de edad, que ocupan el **30%** de la muestra**,** donde las mujeres ocupan el **17%** de las publicaciones y los hombres el **13%** de los posts.

- Los usuarios de **25 a 30** años de edad, que ocupan el **49%** de la muestra, donde las mujeres ocupan el **12%** y los hombres con el **37%** de las publicaciones.
- Los usuarios de **30 a 40** años de edad, que ocupan el **65%** de la muestra**,** donde las mujeres ocupan el **23%** de las publicaciones y los hombres el **42%** de los posts.
- Los usuarios de **40 a 50** años de edad, que ocupan el **50%** de la muestra**,** donde las mujeres ocupan el **29%** de las publicaciones y los hombres el **21%** de los posts.
- Los usuarios de **más de 50** años de edad, que ocupan el **56%** de la muestra**,** donde las mujeres ocupan el **12%** de las publicaciones y los hombres el **44%** de los posts.

- El Grupo Familia se describe con:
  - Los usuarios de **15 a 18** años de edad, que ocupan el **60%** de la muestra**,** donde las mujeres ocupan el **34%** de las publicaciones y los hombres el **26%** de los posts.
  - Los usuarios de **19 a 24** años de edad, que ocupan el **34%** de la muestra**,** donde las mujeres ocupan el **15%** de las publicaciones y los hombres el **19%** de los posts.
  - Los usuarios de **25 a 30** años de edad, que ocupan el **13%** de la muestra, donde las mujeres ocupan el **7%** y los hombres con el **6%** de las publicaciones.
  - Los usuarios de **30 a 40** años de edad, que ocupan el **15%** de la muestra**,** donde las mujeres ocupan el **10%** de las publicaciones y los hombres el **5%** de los posts.
  - Los usuarios de **40 a 50** años de edad, que ocupan el **15%** de la muestra**,** donde las mujeres ocupan el **15%** de las publicaciones y los hombres no comparten este tipo de información.
  - Los usuarios de **más de 50** años de edad, que ocupan el **8%** de la muestra**,** donde las mujeres ocupan el **8%** de las publicaciones y los hombres no comparten este tipo de información.

- El Grupo Reflexión se describe con:
  - Los usuarios de **15 a 18** años de edad, que ocupan el **14%** de la muestra**,** donde las mujeres ocupan el **14%** de las publicaciones y los hombres no comparten este tipo de información.

- Los usuarios de **19 a 24** años de edad, que ocupan el **13%** de la muestra**,** donde las mujeres ocupan el **10%** de las publicaciones y los hombres el **3%** de los posts.
- Los usuarios de **25 a 30** años de edad, que ocupan el **22%** de la muestra, donde las mujeres ocupan el **18%** y los hombres con el **4%** de las publicaciones.
- Los usuarios de **30 a 40** años de edad, que ocupan el **14%** de la muestra**,** donde las mujeres ocupan el **12%** de las publicaciones y los hombres el **2%** de los posts.
- Los usuarios de **40 a 50** años de edad, que ocupan el **22%** de la muestra**,** donde las mujeres ocupan el **18%** de las publicaciones y los hombres el **4%** de los posts.
- Los usuarios de **más de 50** años de edad, que ocupan el **28%** de la muestra**,** donde las mujeres ocupan el **12%** de las publicaciones y los hombres el **16%** de los posts.

- El Grupo Humor se describe con:
  - Los usuarios de **15 a 18** años de edad, que ocupan el **13%** de la muestra**,** donde las mujeres ocupan el **9%** de las publicaciones y los hombres el **4%** de los posts.
  - Los usuarios de **19 a 24** años de edad, que ocupan el **24%** de la muestra**,** donde las mujeres ocupan el **20%** de las publicaciones y los hombres el **4%** de los posts.
  - Los usuarios de **25 a 30** años de edad, que ocupan el **16%** de la muestra, donde las mujeres ocupan el **6%** y los hombres con el **10%** de las publicaciones.
  - Los usuarios de **30 a 40** años de edad, que ocupan el **6%** de la muestra**,** donde las mujeres ocupan el **1%** de las publicaciones y los hombres el **5%** de los posts.
  - Los usuarios de **40 a 50** años de edad, que ocupan el **13%** de la muestra**,** donde las mujeres ocupan el **9%** de las publicaciones y los hombres el **4%** de los posts.
  - Los usuarios de **más de 50** años de edad, que ocupan el **8%** de la muestra**,** donde las mujeres ocupan el **8%** de las publicaciones y los hombres no comparten este tipo de información.

Esta clasificación permite determinar qué contenido es importante y de interés, brinda un aporte significativo para el Trabajo de titulación, por el filtro de información que otorga y la sugerencia o canal de comunicación que aporta para la transmisión de información más directa y personalizada para los usuarios de la provincia de Loja.

4.4 Fase de evaluación e interpretación

Al evaluar los resultados obtenidos es posible determinar que el 56.2% de la muestra recolectada son mujeres y el 43.8% hombres. Donde los usuarios con mayor incidencia de publicaciones en redes sociales son los de 25 a 30 años de edad, tal como es el caso descrito en (Madden & Smith, 2010), donde coinciden que los gestores de redes sociales son los adultos jóvenes de 18 a 29 años de edad, contribuyendo así con el trabajo de titulación realizado para verificar que los resultados no están fuera de enfoque.

Para una mejor comunicación se determinaron 8 tipos de contenido que permiten identificar la preferencia de contenido como en (Barroso et al., 2013). El tipo de contenido se basa en los puntos de mayor relevancia entre las publicaciones encontradas en cada perfil de la red social, donde se destaca la preferencia de contenido determinado como: personal selfies, familia, amistad, reflexión, relación sentimental, música y estado, los mismos que muestran la diversidad en el manejo de datos, permitiendo establecer una estrategia para la transmisión de información, como es el caso político de (López-García, 2016).

La agrupación determinada con la técnica de clustering, mediante el algoritmo mapas de Kohonen, permitió identificar los 8 tipos de contenido de los usuarios de la provincia de Loja. Donde los usuarios tienen interacción baja en publicaciones de música, estado y relación sentimental; interacción media en publicaciones de amistad y humor, mientras que la interacción alta en publicaciones de: personal selfies, reflexión y familia contenido que predomina en los datos de la muestra recolectada.

La clasificación con el algoritmo Kmeans, permitió identificar a varios grupos de los cuales el más adecuado para la información recolectada es de 4. Los 4 grupos identificados son: grupo social, grupo familia, grupo reflexión y grupos humor, los cuales permitieron maximizar las diferencias entre los 8 distintos tipos de contenido y obtener la mejor distribución de los usuarios.

La clasificación por preferencia de contenido para los usuarios de la provincia de Loja y con la implementación de árboles de decisión, es posible mostrar que los usuarios mayores o iguales a 22 años de edad prefieren contenido correspondiente al grupo

humor y grupo familia y para aquellos usuarios mayores a 22 años de edad prefieren contenido del grupo social y grupo reflexión. Determinado así un patrón de consumo de información.

La clasificación por edad identificó que de acuerdo a los 6 grupos de edad, los usuarios se determinan como: los de 15 a 18 años prefieren contenido del grupo social, los de 19 a 24 años prefieren contenido del grupo humor, los de 30 a 40 años prefieren contenido del grupo social, los mayores de 50 años prefieren contenido del grupo reflexión y para los usuarios de 25 a 30 años de edad la preferencia de contenido incluye a los 4 grupos social, familia, reflexión y humor.

En base a los resultados y el análisis correspondiente es posible determinar los perfiles de uso de contenido de los usuarios de la provincia de Loja.

CONCLUSIONES

Con base en la experimentación y la obtención de resultados en los diferentes procesos implementados en cada una de las fases, se puede concluir lo siguiente:

- Los habitantes de la provincia de Loja con la red social más usada Facebook, permiten identificar el interés para una comunicación más directa y personalizada, al compartir contenido de tipo Social, Familiar, Reflexivo y de Humor.

    - El grupo Social tiene mayor presencia de personas que van de los 30 hasta los 40 años de edad, esto refleja que el perfil de su uso de su red social prefiere este tipo de contenido.
    - El grupo Familia se identifica por la preferencia de contenido de los usuarios que van desde los 15 hasta los 18 años de edad.
    - El grupo Reflexión con mayor importancia para los usuarios que tienen más de 50 años de edad.
    - Y el grupo Humor se destaca en las publicaciones de los usuarios que van desde los 19 hasta los 24 años de edad.

- En base a los resultados obtenidos y el análisis correspondiente se destaca las publicaciones del Grupo Social con el 42% de usuarios, que determina la preferencia de contenido y se resalta el uso de imágenes antes que texto, al Grupo Familia con el 24% de usuarios, al Grupo Reflexión con el 18% de usuarios y al Grupo Humor con el 16% de usuarios de la provincia de Loja.

RECOMENDACIONES

A partir de los resultados encontrados en el presente Trabajo de Titulación, se presentan las siguientes recomendaciones:

- En el desarrollo de un Trabajo de Titulación se empiece con la determinación de una metodología acorde a los objetivos y técnicas a resolver, así como se realizó en la implementación de este trabajo con la metodología KDD.

- Revisión periódica de los objetivos y el desarrollo del trabajo para evitar la desorientación y que se cumpla con los plazos planteados descritos en el cronograma.

- Obtener e incluir datos económicos para ver si esto influye o no en el uso de redes sociales y poder determinar patrones de comportamiento relevantes.

- El uso e implementación de varias técnicas de minería de datos, permite evidenciar los resultados esperados.

BIBLIOGRAFIA

Barroso, C., Muñoz de la Luna, Á., & Navarro, E. (2013). Eficacia publicitaria en redes sociales: el caso de Mango en Facebook España, 93–110.

Bernete Francisco. (2010). Usos de las TIC, Relaciones sociales y cambios en la socialización de las y los jóvenes.

Casero-Ripollés, A. (2012). Beyond Newspapers: News Consumption among Young People in the Digital Era. *Comunicar*, *23*(45), 151–158. http://doi.org/10.3916/C39-2012-03-05

El Mundo. (2016). Elecciones. Retrieved from http://www.elmundo.es/cronica/2016/07/03/57779fc0ca4741301d8b4609.html

El Tiempo. (2015). Redes.

Flórez López, R., & Fernández Fernández, J. M. J. M. (2008). *Las redes neuronales artificiales : fundamentos teóricos y aplicaciones prácticas*. Netbiblo.

Gamir Ríos, J. (2016). Blogs, Facebook y Twitter en las Elecciones Generales de 2011. Retrieved from http://roderic.uv.es/bitstream/handle/10550/53623/111615.pdf?sequence=1&isAllowed=y

Hernandez& Abilowo. (2016). Evaluación de modelos para la predicción de la Bolsa.

Hernández, J., Cèsar, O., & Ramírez, F. (2007). T.2 Minería de Datos y Extracción de Conocimiento de Bases de Datos.

Hernández, RAMIREZ, J., QUINTANA, C. R., Orallo, M. Josej. H., Quintana, M. J. R., Ram'\irez, C. F., … others. (2004). *Introducción a la Mineria de Datos*. Pearson Prentice Hall,.

Herrera Castellanos, M. (2016). FORMULA PARA CÁLCULO DE LA MUESTRA POBLACIONES FINITAS.

INEC. (2010). Fascículo Provincia de Loja.

Lara, P., & Duart, J. M. (2005). Gestión de contenidos en el e-learning : acceso y uso de objetos de información como recurso estratégico, *2*, 6–16.

López-García, G. (2016). “Nuevos” y “viejos” liderazgos: la campaña de las elecciones generales españolas de 2015 en Twitter, *29*(293), 149–167. http://doi.org/10.15581/003.29.3.sp.149-167

Madden, M., & Smith, A. (2010). Reputation Management and Social Media Summary of Findings. Retrieved from http://www.pewinternet.org/files/old-media/Files/Reports/2010/PIP_Reputation_Management_with_topline.pdf

Merelo, J, J. (2004). Tutorial: mapas organizativos de Kohonen. Retrieved February 5, 2017, from http://geneura.ugr.es/~jmerelo/tutoriales/bioinfo/Kohonen.html

Mezrich, B., & Vilà Vernis, R. (2010). *Multimillonarios por accidente : el nacimiento de Facebook : una historia de sexo, dinero, talento y traición*. Alienta.

Microsoft. (2016). Conceptos de minería de datos. Retrieved June 6, 2017, from https://msdn.microsoft.com/es-ES/library/ms174949(v=sql.120).aspx

Peña, A. la, & Parra, T. (2016). ALGORITMOS GENÉTICOS.

Ramírez, A. Y. (2007). Técnicas de Minería de Datos Aplicadas a la Construcción de Modelos de Score Crediticio :, 1–10.

Red Internacional. (2016). Medios Digitales.

Resina, J. (2010). Ciberpolítica , redes sociales y nuevas movilizaciones en España : el impacto digital en los procesos de deliberación y participación ciudadana Cyberpolitics , Social Networks and New Mobilizations in Spain : Digital Impact on the Processes of Deliberatio, 143–164.

Rodríguez, O. (2014). Lección Nº4 Calibración y Selección de Modelos - LAPLACE - YouTube. Retrieved June 4, 2017, from https://www.youtube.com/watch?v=UwX4Ta78J0U

Rui Xu, D. W. (2009). Clustering. Retrieved February 5, 2017, from https://books.google.com.ec/books?hl=es&lr=&id=kYC3YCyl_tkC&oi=fnd&pg=PR5&dq=wunsch+2009+clustering&ots=qid7zJec0F&sig=MmunyAFltGABTxkyXljp54UpDA4#v=onepage&q=wunsch 2009 clustering&f=false

Ruiz del Olmo, F., & Bustos Díaz, J. (2016). *Del tweet a la fotografía, la evolución de la comunicación política en Twitter hacia la imagen. El caso del debate del estado de la nación en España (2015)*. La Laguna, Tenerife. http://doi.org/10.4185/RLCS-2016-1086

Serrano Santoyo Evelio Martínez Mtz, A. (2003). LA BRECHA DIGITAL Mitos y

Realidades. *México Teléfono*, (686), 552–1056. Retrieved from http://www.uabc.mx/

Stillo, M. (2012). Medios de comunicación y Democracia.

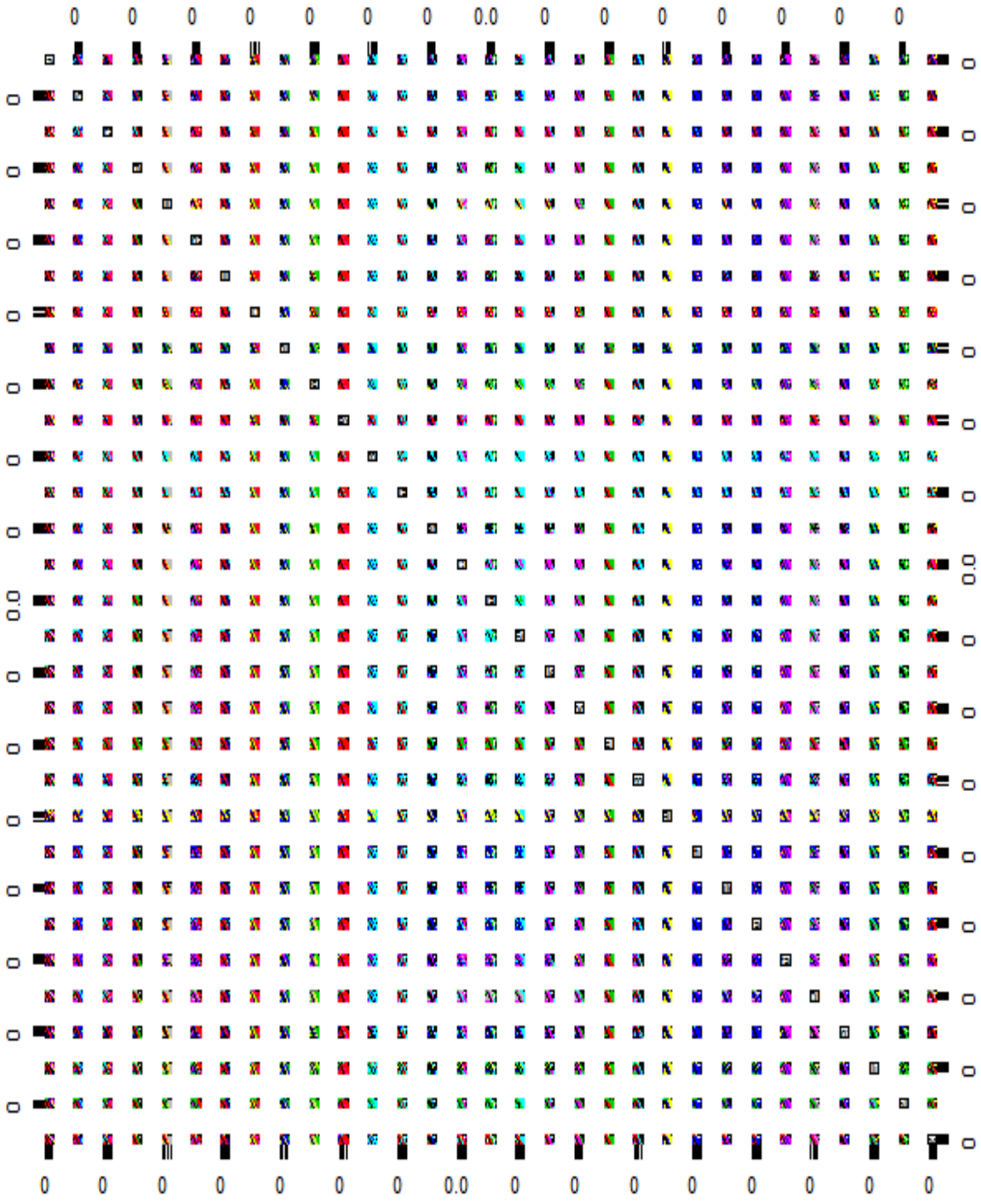
Valencia Z., G. A. (2012). MINERÍA DE DATOS LA MINERÍA DE DATOS COMO HERRAMIENTA PARA LA TOMA DE DECISIONES ESTRATÉGICAS. Retrieved from http://gustavovalencia.net/app/webroot/img/Documents/BI/Actividades/001/Articulo DM.pdf

Weka. (2011). Software de minera de datos en Java. Retrieved January 5, 2017, from http://www.cs.waikato.ac.nz/ml/weka/

ANEXOS

En las fases descritas para el desarrollo del Trabajo de Titulación, la recopilación de información permitió identificar 31 categorías que al aplicarle un algoritmo de clasificación y como se muestra en la Figura no es posible describir la preferencia de contenido de los usuarios.

- **Anexo I CLUSTERING ALGORITMO KMEANS 31 CATEGORIAS**



Luego de la selección de datos, se procedió a una clasificación con los ocho tipos de contenido, variando los grupos de clasificación para una mejor interpretación entre los cuales se menciona:

- **Anexo II CLASIFICACION DE 2**

Centros de clústeres finales

	Clúster	
	1	2
personal selfies	1	8
Reflexión	3	1
Familia	3	2
Amistad	2	4
Humor	2	1
Música	1	0
Estado	1	1
relación sentimental	1	1

ANOVA

	Clúster		Error			
	Media cuadrática	Gl	Media cuadrática	gl	F	Sig.
personal_selfie	4937,464	1	5,357	500	921,688	,000
reflexion	403,429	1	9,319	500	43,292	,000
familia	40,627	1	9,333	500	4,353	,037
amistad	505,137	1	6,448	500	78,340	,000
humor	125,397	1	7,109	500	17,640	,000
musica	35,721	1	3,771	500	9,472	,002
estado	16,858	1	2,620	500	6,434	,011
relacion_s	7,715	1	3,508	500	2,199	,139

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

Número de casos en cada clúster

Clúster	1	357,000
	2	145,000
Válidos		502,000
Perdidos		,000

- **Anexo III CLASIFICACION DE 3**

Centros de clústeres finales

	Clúster		
	1	2	3
personal selfies	1	2	9
reflexión	6	1	1
familia	1	4	2
amistad	1	2	3
humor	4	1	1
música	1	1	1
estado	1	1	1
relación sentimental	0	1	1

Número de casos en cada clúster

Clúster	1	148,000
	2	228,000
	3	126,000
Válidos		502,000
Perdidos		,000

- **Anexo IV CLASIFICACION DE 4 GRUPOS**

Centros de clústeres finales

	Clúster			
	1	2	3	4
personal selfies	1	1	2	9
Reflexión	1	8	3	1
Familia	4	2	1	2
Amistad	2	1	1	3
Humor	1	2	7	1
Música	1	1	1	1
Estado	1	1	1	1
relación sentimental	1	0	0	1

ANOVA

	Clúster		Error			
	Media cuadrática	gl	Media cuadrática	gl	F	Sig.
personal selfies	1755,607	3	4,717	498	372,177	,000
Reflexión	1046,736	3	3,861	498	271,134	,000
Familia	171,612	3	8,418	498	20,386	,000
Amistad	99,500	3	6,889	498	14,444	,000
Humor	729,902	3	2,992	498	243,941	,000
Música	13,795	3	3,775	498	3,654	,013
Estado	10,583	3	2,601	498	4,070	,007
relación sentimental	17,569	3	3,432	498	5,119	,002

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

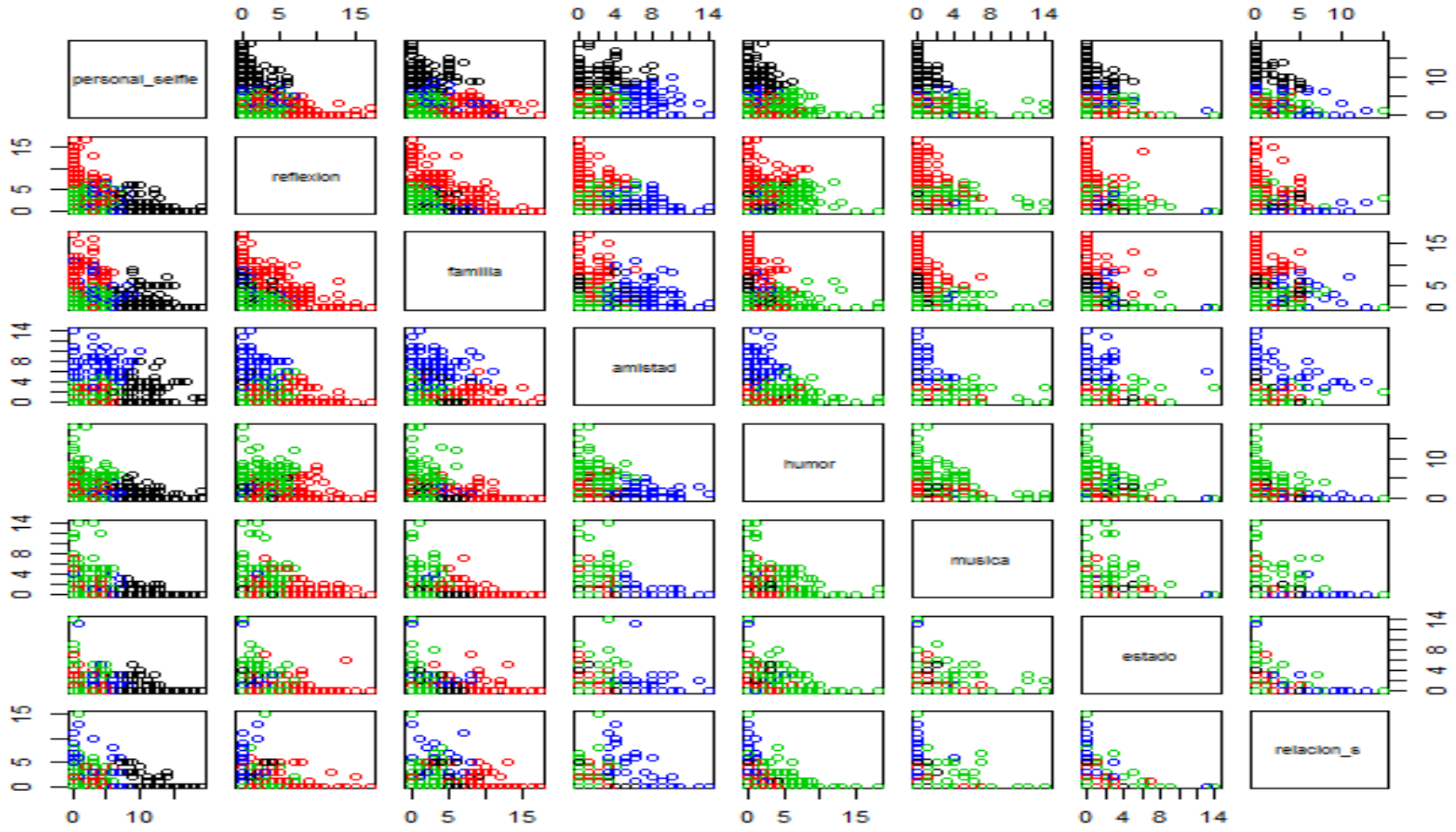
Número de casos en cada clúster

Clúster	1	213,000
	2	90,000
	3	79,000
	4	120,000
Válidos		502,000
Perdidos		,000

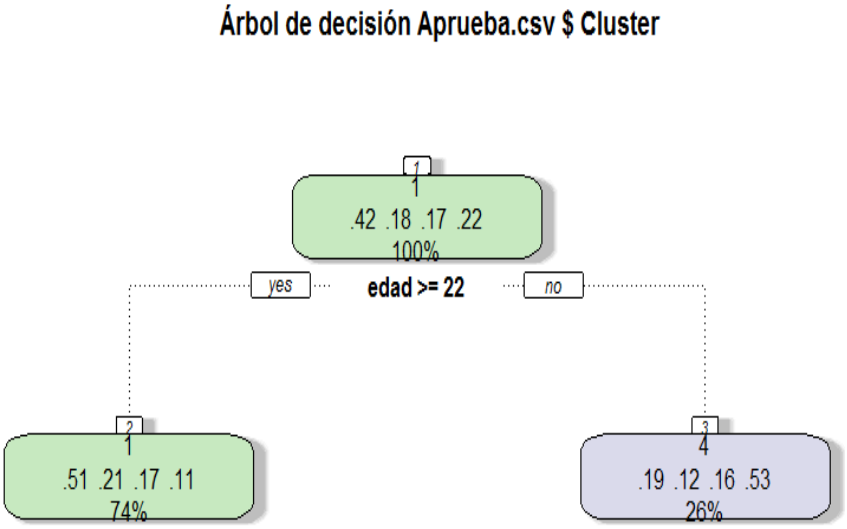
Con la clasificación de 4 grupos se identifica una mejor distribución de datos por lo cual es la que determina como adecuada para el tratamiento y análisis de minería.

Con ayuda de la herramienta R Studio y la aplicación del algoritmo K-means con los 4 grupos de contenido determinados y como se muestra en la siguiente Figura la distribución de datos, permite una mejor interpretación de información.

- **Anexo V CLASIFICACION POR TIPO DE CONTENIDO**



- **Anexo VI ARBOL DE DECISION**



Modelo del Árbol de decisión.

Input: edad
Number of observations: 502

```
crs$rpart <- rpart(Cluster ~ .,
    data=crs$dataset[crs$train, c(crs$input, crs$target)],
    method="class",
    parms=list(split="information"),
    control=rpart.control(usesurrogate=0,
        maxsurrogate=0))
```

Árbol como reglas:

```
 Rule number: 3 [Cluster=4 cover=93 (26%) prob=49.00]
   edad< 22.5
 Rule number: 2 [Cluster=1 cover=258 (74%) prob=29.00]
   edad>=22.5

1) edad <= 22; criterion = 1, statistic = 46.317
  2) edad <= 17; criterion = 0.958, statistic = 4.14
    3)*  weights = 39
  2) edad > 17
    4)*  weights = 96
1) edad > 22
  5) edad <= 25; criterion = 0.951, statistic = 3.859
    6)*  weights = 81
  5) edad > 25
    7)*  weights = 286
```

Anexo VII ESTADISTICAS

Casosclúster4		Media	Desviación estándar	N válido (por lista)	
				No ponderados	Ponderados
1	personal_selfie	1,49	1,659	213	213,000
	reflexion	1,21	1,498	213	213,000
	familia	3,50	3,883	213	213,000
	amistad	2,44	2,980	213	213,000
	humor	,75	1,177	213	213,000
	musica	1,09	2,564	213	213,000
	estado	1,07	2,049	213	213,000
	relacion_s	1,12	2,393	213	213,000
2	personal_selfie	,94	1,425	90	90,000
	reflexion	7,87	2,957	90	90,000
	familia	1,66	1,961	90	90,000
	amistad	1,37	1,875	90	90,000
	humor	1,74	1,833	90	90,000
	musica	,60	1,140	90	90,000
	estado	,54	1,113	90	90,000
	relacion_s	,33	1,028	90	90,000
3	personal_selfie	2,09	2,101	79	79,000
	reflexion	3,06	2,078	79	79,000
	familia	,94	1,471	79	79,000
	amistad	1,37	1,673	79	79,000
	humor	6,67	2,960	79	79,000
	musica	1,24	1,689	79	79,000
	estado	,84	1,454	79	79,000
	relacion_s	,43	1,129	79	79,000
4	personal_selfie	9,04	3,229	120	120,000
	reflexion	1,24	1,670	120	120,000
	familia	1,84	2,017	120	120,000
	amistad	3,41	2,923	120	120,000
	humor	,95	1,340	120	120,000
	musica	,53	1,115	120	120,000
	estado	,51	1,045	120	120,000
	relacion_s	,69	1,592	120	120,000
Total	personal_selfie	3,29	3,899	502	502,000

reflexion	2,70	3,179	502	502,000
familia	2,37	3,065	502	502,000
amistad	2,31	2,728	502	502,000
humor	1,91	2,710	502	502,000
musica	,89	1,958	502	502,000
estado	,80	1,627	502	502,000
relacion_s	,77	1,875	502	502,000

Funciones en centroides de grupo

	Función		
Casosclúster4	1	2	3
1	-,094	-1,428	-,391
2	-2,162	,548	1,731
3	-1,268	1,923	-1,689
4	2,624	,859	,507

Las funciones discriminantes canónicas sin estandarizar se han evaluado en medias de grupos

- **Anexo VII RELACIÓN DE VARIABLES**

	FAMILIA		HUMOR		REFLEXION		SOCIAL		Usuarios	Genero
Edad	**Usuarios**	**Genero**	**Usuarios**	**Genero**	**Usuarios**	**Genero**	**Usuarios**	**Genero**		
15-18										
F	58%	25	14%	6	19%	8	9%	4	100,00%	43
M	65%	11	12%	2	0%		24%	4	100,00%	17
Total 15-18	**60%**	**36**	**13%**	**8**	**13%**	**8**	**13%**	**8**	**100,00%**	**60**
F	31%	27	30%	26	15%	13	24%	21	100,00%	87
M	39%	15	11%	4	8%	3	42%	16	100,00%	38
Total 19-24	**34%**	**42**	**24%**	**30**	**13%**	**16**	**30%**	**37**	**100,00%**	**125**
F	13%	13	11%	11	32%	33	44%	45	100,00%	102
M	13%	10	24%	18	8%	6	55%	42	100,00%	76
Total 25-30	**13%**	**23**	**16%**	**29**	**22%**	**39**	**49%**	**87**	**100,00%**	**178**
F	18%	7	3%	1	20%	8	60%	24	100,00%	40
M	12%	3	12%	3	4%	1	73%	19	100,00%	26
Total 30-40	**15%**	**10**	**6%**	**4**	**14%**	**9**	**65%**	**43**	**100,00%**	**66**
F	21%	7	6%	2	26%	9	47%	16	100,00%	34
M	0%		29%	4	14%	2	57%	8	100,00%	14
Total 40-50	**15%**	**7**	**13%**	**6**	**23%**	**11**	**50%**	**24**	**100,00%**	**48**
F	14%	2	14%	2	21%	3	50%	7	100,00%	14
M	0%		0%		36%	4	64%	7	100,00%	11
Total 50-mas	**8%**	**2**	**8%**	**2**	**28%**	**7**	**56%**	**14**	**100,00%**	**25**
Total, general	**24%**	**120**	**16%**	**79**	**18%**	**90**	**42%**	**213**	**100,00%**	**502**