



UNIVERSIDAD CÉSAR VALLEJO

**FACULTAD DE INGENIERÍA Y ARQUITECTURA
ESCUELA PROFESIONAL DE INGENIERÍA DE
SISTEMAS**

“Identificación de patrones de consumo de Servicios Turísticos en redes sociales aplicando minería de datos, Piura 2022”

**TESIS PARA OBTENER EL TÍTULO PROFESIONAL DE:
INGENIERO DE SISTEMAS**

AUTOR:

Ramos Morales, Anthony (orcid.org/[0000-0001-9594-8731](https://orcid.org/0000-0001-9594-8731))

ASESOR:

Mg. Agurto Marchan, Winner (orcid.org/[0000-0002-0396-9349](https://orcid.org/0000-0002-0396-9349))

LÍNEA DE INVESTIGACIÓN:

Sistema de Información y Comunicaciones

LÍNEA DE RESPONSABILIDAD SOCIAL UNIVERSITARIA

Desarrollo económico, empleo y emprendimiento

Piura – Perú

2022

DEDICATORIA

El presente trabajo está dedicado a mis padres, por el empeño y trabajo durante estos años, para que pueda lograr mis metas y mis objetivos. Sin ellos, no hubiese podido conseguir lo que hasta ahora. A mi esposa y a mi hija por el amor y el apoyo incondicional y por ser la inspiración para seguir adelante. A mis hermanos por acompañarme y animarme en todo momento a ser mejor día a día.

AGRADECIMIENTO

Agradezco a Dios todopoderoso, por cuidarme siempre durante todo este recorrido.

A mis padres por el apoyo y la confianza brindada para culminar mi carrera profesional.

A mi esposa e hija por su apoyo incondicional, por ser mi fortaleza en los momentos difíciles.

A mis hermanos por siempre estar allí cuando los necesito.

Al docente del curso por sus sugerencias y consejos que hicieron posible concluir la presente investigación.

A todas las personas que contribuyeron, dedicando su tiempo y cooperación dentro de lo personal y profesional durante esta investigación.

ÍNDICE

I. INTRODUCCIÓN	2
II. MARCO TEÓRICO	5
III. METODOLOGÍA.....	14
3.1. Tipo y diseño de investigación:	14
Tipo de investigación:	14
3.2. Variables y operacionalización:.....	15
3.3. Población, muestra y muestreo:.....	15
3.4. Técnicas e instrumentos de recolección de datos:.....	16
3.5. Procedimientos:	16
3.6. Método de análisis de datos:.....	17
3.7. Aspectos éticos:	17
IV. RESULTADOS:	18
V. DISCUSIÓN:.....	22
VI. CONCLUSIONES.....	25
VII. RECOMENDACIONES	26
REFERENCIAS BIBLIOGRÁFICAS	28
ANEXOS:	32

ÍNDICE DE TABLAS

Tabla 1 - Consumo de tipo de publicaciones según Mes	18
Tabla 2 - Grupo de Consumo según Servicio.....	20

ÍNDICE DE FIGURAS

Figura 1 - Componentes del aprendizaje automático	8
Figura 2 - Ejemplo gráfico de dispersión con varios Clusters.....	10
Figura 3 - Técnicas de minería de datos más utilizadas	11
Figura 4 - Técnicas de minería de datos	13
Figura 5 - Cálculo del número óptimo de Clusters	19
Figura 6 - Grupo de Servicio según Clúster	20
Figura 7 - Grupo de Consumo de Servicios según Temporada	21

RESUMEN

Hoy en día el uso de las nuevas herramientas tecnológicas; como la minería de datos, están impactando de manera positiva en los diferentes ámbitos de la sociedad, para lo cual la presente investigación tuvo como objetivo identificar patrones de consumo de servicios turísticos en las redes sociales utilizando técnicas de minería de datos. El sustento teórico fue en base a minería de datos, y sus técnicas tales como Clustering y asociación dentro de las más destacadas. También se hizo uso de teorías de redes sociales, Turismo y paquetes turísticos. Así mismo fue desarrollada en base a una investigación de tipo aplicada, con un nivel de profundidad descriptivo y la naturaleza de los datos tratados fueron de tipo cuantitativo, teniendo como población y muestra usuarios de la red social Facebook que interactuaron con el #viajesPerú. Se obtuvo como principales resultados la identificación de diversos patrones de consumo de servicios turísticos en la red social Facebook haciendo uso de técnicas como Clustering y asociación. Llegando a la conclusión que la minería de datos permite obtener información de tipo cualitativos y convertirlos en datos numéricos, proporcionando información a diferentes empresas para mejorar la capacidad de toma de decisiones y de planificación de estrategias.

Palabras clave: Minería de datos, servicios turísticos, técnicas de minería de datos.

ABSTRACT

Nowadays the use of new technological tools such as data-mining is having a positive impact on different areas of society, for which this study aimed to identify patterns of consumption of tourism services in social networks using data mining techniques. The theoretical support was based on data mining, and its techniques such as clustering and association among the most prominent. Theories of social networks, tourism and tourism packages were also used. Likewise, it was developed based on an applied type of research, with a descriptive level of depth and the nature of the data treated was quantitative, having users of the social network Facebook who interacted with #viajesPerú as population and sample. The main results obtained were the identification of various patterns of consumption of tourism services in the social network Facebook using techniques such as clustering and association. It was concluded that data-mining enables the obtaining of qualitative information and its conversion into numerical data, providing information for different companies to improve their decision-making and strategy-planning capabilities.

Keywords: Data-mining, tourism services, data-mining techniques.

I. INTRODUCCIÓN

El avance tecnológico y el masivo uso de redes sociales están obligando a las empresas adaptarse rápidamente a las nuevas innovaciones y requerimientos de sus consumidores. Con ello las TIC (Tecnologías de Información y Comunicaciones) están logrando convertirse en uno de los pilares bases de nuestra sociedad, siendo un elemento muy importante dentro de una organización, ayudándolas no solo a adaptarse a un entorno dinámico y variado, sino también aportando ventajas competitivas para lograr avances significativos (Insightsoftware, 2019).

En este sentido (Aggarwal, 2018) define las redes sociales como una estructura que permite la difusión de información y que está compuesta por actores o entidades que facilitan el intercambio de contenido que se puede llegar a compartir, generando tendencias y patrones interesantes. Por su lado (Donald Jense, y otros, 2002) indican que la amistad entre personas, vínculos entre páginas, lugares, cosas son parte de este conjunto que compone las redes sociales.

Así mismo (Guerrero González, y otros, 2014) mencionan que la tecnología está generando un gran impacto e influencia positiva dentro del sector turístico. Destacan los avances que se han alcanzado en infraestructura, hospedaje, transporte y sobre todo recalcan las facilidades que se han obtenido para el manejo de información, búsqueda de productos turísticos y las reservas de estos. En este contexto se puede destacar que hoy en día los consumidores de servicios turísticos han cambiado la forma en la que viajan; como realizan cada una de las búsquedas para obtener información antes y durante el viaje (elección del lugar de destino, el hospedaje, los principales sitios que deben visitar, etc.) y la forma en la cual comparten dicha información (fotos, experiencias, recomendaciones, etc.) (Aušrinė, 2021). Estos usuarios además de buscar información son creadores de contenido, ya que están en constante interacción realizando diferentes acciones en la red (Alhajj , y otros, 2018).

De acuerdo con el portal (Statista, 2021) indica que el sector turismo contribuyó con 4,7 billones de dólares al PBI mundial en el año 2020. Así mismo la (OMT, 2020) destaca la fortaleza y crecimiento del sector en la última década. De la misma manera indica que se vienen generando cambios positivos en los sectores rurales, generando oportunidades de trabajo y sobre todo impulsando el crecimiento económico en el sector. Por su lado (The World Bank Group, 2017) recalca que de manera particular el turismo destaca con mayor importancia en los países de bajos ingresos, los cuales clasifica en un total de 48. Y así mismo señala que en el 2015 estos países recibieron cerca de 29 millones de turistas internacionales.

En el Perú el sector turístico ha favorecido de manera significativa al crecimiento de la economía. Según (Bucaram, y otros, 2019) indican que el turismo se incrementó en una media anual del 9,0% entre los años 2010 y 2019. Teniendo una afluencia de 4,4 millones de turistas donde se alcanzaron los 4789 millones de dólares en el 2019. Por lo que el diario (El Peruano, 2022) resalta que el turismo contribuyó con el 2% al PBI del país en el año 2021; destacando que antes de pandemia este sector contribuyó con el 3,9%. Todo esto se debe sin duda a las diversas ventajas competitivas que tiene el país en la región por los diversos atractivos geográficos, históricos y culturales (PROMPERÚ, 2016). Por otro lado, yendo más a fondo a nivel regional (MINCETUR, 2018) , destaca a Piura dentro de las 13 regiones con mayor competitividad a nivel turístico; destacando con el 6,6% del total de viajes realizados a nivel interno. Así mismo (MINCETUR, 2021) indica que según datos de INEI se incrementó la actividad en este sector en un 11,7% durante el año 2021. Es importante mencionar las diferentes iniciativas innovadoras que se están implementando para involucrar a los pobladores de las zonas turísticas, como medida para combatir la pobreza en dichas zonas (Figuroa Pinedo, 2019).

A pesar de todo lo mencionado anteriormente y de todo lo que se viene logrando en este sector; hoy en día una de las principales problemáticas que ostentan algunas organizaciones, es la escasa información sobre las ventajas y beneficios que aporta la Minería de Datos a la búsqueda, comprensión y extracción de grandes cantidades de información importante disponible y de forma gratuita; que contribuirán en la mejora y crecimiento de su empresa; pero sobre todo que

ayudaran a la observación y análisis para la toma de decisiones que derivan de este fenómeno.

“La minería de datos o data Mining se define como el conjunto de técnicas de análisis y herramientas que facilitan comprender grandes cantidades de datos, teniendo como finalidad revelar comportamientos y/o patrones repetitivos que aclaren la conducta de los datos en cuestión” (Rivas, 2018).

El crecimiento del interés de los consumidores sobre los servicios turísticos hoy en día, obligan a las empresas del sector a contar con la información adecuada y eficaz de cada consumidor, y como sabemos que cada uno tiene diferentes necesidades y requerimientos para un producto o servicio; y el no tener o definir patrones en cuanto a la información importante de los consumidores, impedirá adelantarse o predecir nuevas oportunidades, tendencias y preferencias que estos requieran, y sobre todo que contribuyan a ofrecer un mejor servicio. Entonces, ¿Cómo identificar patrones de consumo de servicios turísticos en las redes sociales mediante la minería de datos?

Como objetivo principal se identificará patrones de consumo de servicios turísticos en las redes sociales utilizando las técnicas de minería de datos. Así como los objetivos específicos los cuales son: establecer las diferentes técnicas de minería de datos que se van a utilizar y evaluar ciertos patrones para determinar la futura toma de decisiones.

En este trabajo de investigación se obtendrá un modelo para identificar los patrones de consumo que ciertos usuarios presenten en las redes sociales, usando técnicas de minería de datos; las cuales nos brindará datos importantes y segmentados para identificar las características de consumo de estos, y con esto las organizaciones puedan tomar mejores decisiones para mejorar los recursos que ellos brindan.

II. MARCO TEÓRICO

Respecto a la investigación de trabajos tanto nacionales e internacionales relacionados con este trabajo de investigación, se encontraron los siguientes:

(De Lucca, 2017) en su investigación sobre: “Minería de datos y el turismo”; se propuso como objetivo analizar la utilidad del Data Mining como una herramienta inteligente para que esta potencie la imagen de la marca argentina como un destino turístico internacional. La metodología fue de tipo exploratoria. Concluyendo, que gracias al avance de las diversas tecnologías de información ha impactado en la demanda y la oferta del sector turístico, los nuevos turistas buscan vivir experiencias y estas compartirlas por redes sociales, la imagen turística de un destino tiene una gran repercusión por el contenido generado.

(Alania Ricaldi, 2018) en su investigación titulada: “Aplicación de la Minería de Datos para pronosticar el abandono de sus estudiantes de la facultad de Ingeniería de la Universidad Alcides Carrión” teniendo como objetivo: pronosticar el abandono de sus estudiantes utilizando técnicas de minería de datos para la universidad en mención. Se utilizó como grupo objetivo de estudio a estudiantes de la Escuela de Sistemas de los años 2013 al 2017. Siendo los resultados que existe una influencia entre el abandono y el promedio de sus notas.

(Uriarte del Águila, 2018) en su investigación sobre: “Minería de datos para optar por la mejor toma de decisiones para la gestión al cliente de telefonía de la zona de Tarapoto”, se pretendió demostrar en qué medida el uso de esta herramienta podría contribuir con la mejor toma de decisiones a la hora de brindar el servicio hacia sus clientes. Después de procesar los datos correspondientes, se llegó a la conclusión que con el uso de esta herramienta se ha mejorado sustancialmente el proceso que se da en la toma de decisiones en dicha área. Asimismo, se recomendó el uso de esta herramienta para otras áreas dentro de la organización de Telefonía.

(Córdova Valdivia, y otros, 2018) en su trabajo titulado “Aplicación de la Minería de Datos para pronosticar el riesgo de morosidad en estudiantes de la UAP”, tiene como objetivo: determinar la medida en la que está técnica influenciará en el pronóstico de riesgo de morosidad en estudiantes de la UAP en el año 2016. Para esta investigación se realizó con el uso de la técnica de observación directa y la aplicación de encuestas y según los resultados se comprobó que el modelo predictivo tiene un alto porcentaje de aseveración.

(Arévalo Rodríguez, y otros, 2016) en su investigación hace mención sobre la Identificación de patrones de consumo de los servicios turísticos para el sector hotelero en la ciudad de Trujillo, basadas en la información de las fanpages de las empresas hoteleras de la ciudad y como muestra el fanpage de la empresa Kallpa Hotel, utilizando técnicas de agrupamiento o Clustering y reglas de asociación. Llegando a concluir que el patrón de comportamiento de los clientes en los primeros meses del análisis es constante y la mitad del año tienen tendencia a aumentar.

(Jaramillo Valbuena, y otros, 2015) a través de su publicación en la Red de Revistas Científicas sobre Minería de datos como herramienta para los streams en redes sociales, en Buenos Aires, Argentina; nos menciona de cuán beneficioso es la Minería de datos para la Bibliotecología ya que obtiene conocimiento en tiempo real, identificando fortalezas y/o debilidades mediante el análisis de opiniones.

Para poder entender el presente trabajo de investigación a continuación, se exponen los aspectos teóricos con mayor relevancia.

Tal es el caso del turismo que se define como el movimiento temporal y voluntario un individuo o de un grupo de personas (Fuller, 2009). Siendo esta una de las actividades con gran notoriedad y protagonismo en la actualidad, debido a que es muy diversa y multidisciplinaria; ya que engloba espacios de gran importancia para el desarrollo de cualquier país en el mundo (Figuroa Pinedo, 2018). Así mismo este servicio se define como el conjunto de diversas actividades perfectamente diferenciadas, pero al mismo tiempo relacionadas entre sí, generando un movimiento de manera coordinada y armónica con la finalidad de poder dar respuesta a las diferentes exigencias de los distintos servicios

planteados por la estructura socioeconómica de un determinado flujo turístico (Guerrero González, y otros, 2014).

Según (Antonietti, 2017) dentro de los principales servicios turísticos podemos encontrar el **alojamiento**, que son establecimientos que proporciona de un lugar para albergarse y este además puede ofrecer otros servicios complementarios. Así mismo se menciona las **atracciones** como los conjuntos de servicios que no se incluyen en la actividad principal que proporciona el alojamiento turístico. Si no más bien, son servicios turísticos complementarios. Que pueden ser parques temáticos dedicados a proporcionar diferentes actividades y/o servicios con el fin de brindar entretenimiento y diversión a sus consumidores. Pueden ser de índole deportivo, enfoque medioambiental, de tipo cultural, etc.

Además, se indica que los **servicios complementarios** son todos los que como su nombre nos revela, permiten complementar las diferentes necesidades de los turistas en cualquier parte o zona turística. Por último, se tiene a los **restaurantes** como los lugares que en su gran mayoría se bonifica por el consumo de comida y bebida dentro del establecimiento. En la actualidad hay una gran cantidad de variedades en la modalidad del servicio y los diversos tipos de cocina.

Por su parte (Oliveira, 2013) indica que la interacción constante entre las TICs y el sector turismo están generando nuevas condiciones dentro de este mercado. El contenido que se genera de manera online, a través de medios como las redes sociales; son un componente clave ya que proporciona información para mejorar las diferentes estrategias y sobre todo generar competitividad en el entorno; elevando los estándares de calidad tanto en los servicios como en los diferentes lugares de destino.

Por otro lado, la minería de datos se define como el proceso para clasificar grandes volúmenes de datos utilizando diferentes técnicas o herramientas, con la finalidad de establecer patrones, comportamientos, tendencias o características repetitivas, con las cuales se pueda comprender la información obtenida contribuyendo en el proceso de tomar mejores decisiones en un entorno específico. Con respecto a ello (Microsoft, 2014) por su parte destaca como principales características de la minería de datos la predicción y el aprendizaje

automático. Los cuales son de vital importancia para predecir tendencias, generar reglas y recomendaciones, identificar patrones; pero sobre todo generar y/o obtener nueva información.

Así mismo (Han, y otros, 2011) afirman que la minería de datos a dado y seguirá dando grandes pasos que permite convertir un gran conjunto de datos en conocimiento valioso y organizado. Por su parte (Palma, y otros, 2009) contribuyen a esta definición afirmando que la minería de datos permite la preparación, el sondeo y la exploración de datos para extraer toda la información oculta que en ellas exista, de modo que esos pequeños descubrimientos pueden llegar a generar un alto impacto.

(Witten, y otros, 2011) indican que la minería de datos es un tema donde el aprendizaje se da en base a la práctica; en esencia dicho aprendizaje se basa en la adquisición del conocimiento y la capacidad para utilizarlo. Para ello se destacan tres componentes fundamentales donde el primero son los **datos obtenidos**; y se dice que estos se pueden recolectar de manera manual y automática. Luego se destacan las **características**; que son las variables o patrones identificados que contribuyen a construir un conjunto de datos representativos (Viera, y otros, 2009). Y finalmente se encuentra el **algoritmo**; que es la manera en la que se puede resolver el problema.

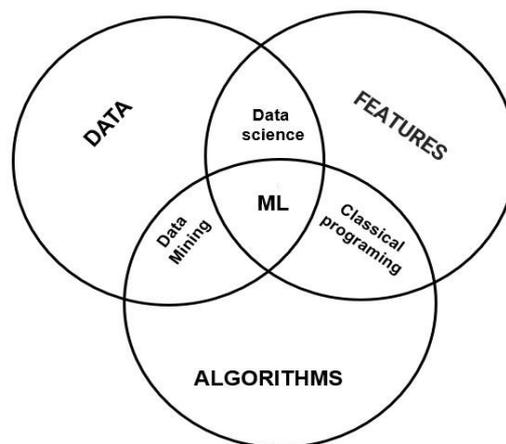


Figura 1 - Componentes del aprendizaje automático

Según (Belinchón Monjas, 2015) estos algoritmos se clasifican en *Forescasting* o *predicción*, mediante el cual permite predecir resultados a partir del análisis estadístico de tendencias obtenidas, para definir estrategias a largo plazo.

Luego se tiene el algoritmo *supervisados* o *predictivos* que permiten predecir cierta información o un conjunto de ellas, con datos que se llegaron por la razón a partir de otros datos conocidos. Finalmente se encuentra el *algoritmo no supervisado* y este último permite encontrar comportamientos y preferencias en los datos manipulados.

Así mismo dentro de las diferentes técnicas y algoritmos que van a permitir realizar diversos análisis, se debe tener en claro cuales se deben elegir. En este sentido según (Microsoft, 2022) se destacan cuatro aspectos fundamentales donde; el primero consiste en la **determinación de los objetivos**; aquí se debe tener en claro, qué objetivos se desea alcanzar con el uso de estas técnicas a partir de las necesidades del cliente. Luego se encuentra el **procesamiento de los datos**; que se da con la clasificación y transformación de la información.

Después, aparece la **definición del modelo**; el cual empieza con la observación estadística de los datos, para luego desarrollar el diseño de la visualización de estos con la ayuda de algunos algoritmos. Finalmente, el **análisis de los resultados**; se da con la verificación de la coherencia del resultado y los datos obtenidos. Es aquí donde el cliente ve si los resultados son útiles y beneficiosos para realizar sus objetivos iniciales.

Siguiendo en esta línea a continuación, se procede a detallar más a fondo a cerca de las diferentes técnicas más representativas de minería de datos donde se tiene a los **Árboles de decisión**; los cuales son modelos de predicción a partir del cual se construyen diagramas lógicos en base a la utilización de un algoritmo que, empezando de un nodo (atributos de entrada), extiende ramas (decisiones) y luego se determina el resultado. Esto sirve para poder representar y calificar una serie de condiciones que surgen en sucesión para resolver un problema.

Luego tenemos las **Redes Neuronales** que según (Hamilton, 2014) son poderosas herramientas de aprendizaje automático que permite analizar grandes volúmenes de datos y de esta manera dan sentido a entornos complejos. Por su parte (Loy, 2019) indica que son modelos o paradigmas de aprendizaje con un procesamiento automático que, en cierta medida, imitan las conexiones entre neuronas; es decir, son un conjunto de nodos o bien llamados neuronas artificiales conectadas transmitiendo señales entre ellas, desde la entrada hasta generar una salida (Oladipupo, 2014).

La siguiente técnica para destacar es el **Clustering** o también llamada de agrupamiento; son algoritmos clasificados como no supervisados, que permiten descubrir grupos o estructuras que se expresan por sí mismas y no depende de alguna circunstancia dentro de los datos estudiados. Este algoritmo considera criterios de cercanía o similitud para realizar su análisis. Dentro de los principales algoritmos que podemos encontrar en esta técnica son los de **K-means**, **agrupamiento jerárquico** y **DBSCAN**. Siendo **K-means** uno de los algoritmos mas popular, ya que principalmente lo que hace es dividir el espacio de datos en K grupos diferentes. El cual inicia con centros de conglomerados de K seleccionados al azar. Posteriormente los centros de los grupos o conglomerados se vuelven a calcular y estos puntos de datos se reasignan a los grupos con mayor cercanía. Dicho proceso se realiza tantas veces hasta que los centros de cada grupo dejen de moverse.

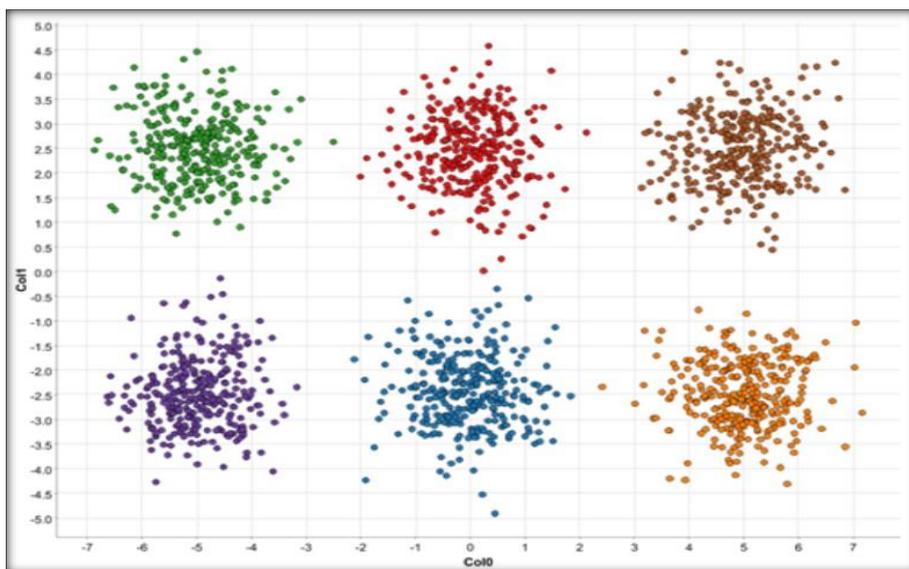


Figura 2 - Ejemplo gráfico de dispersión con varios Clusters

Por su lado la agrupación de Jerarquía de grupos funciona a través de la interconexión iterativa de los puntos mas cercanos para formar grupos. Mientras que DBSCAN, significa Agrupación espacial basada en la densidad de aplicaciones con ruido.

Según (KDnuggets, 2019) en la encuesta realizada a sus lectores en el año 2019, se observa que la técnica de Clustering se encuentra dentro del top 3 de las 10 técnicas más utilizadas para la descripción de los datos. Sumado a esto, la definición teórica de esta técnica es la que más se acopló al presente estudio; ya que permitió describir los grupos con más cercanía o solicitud para el análisis.

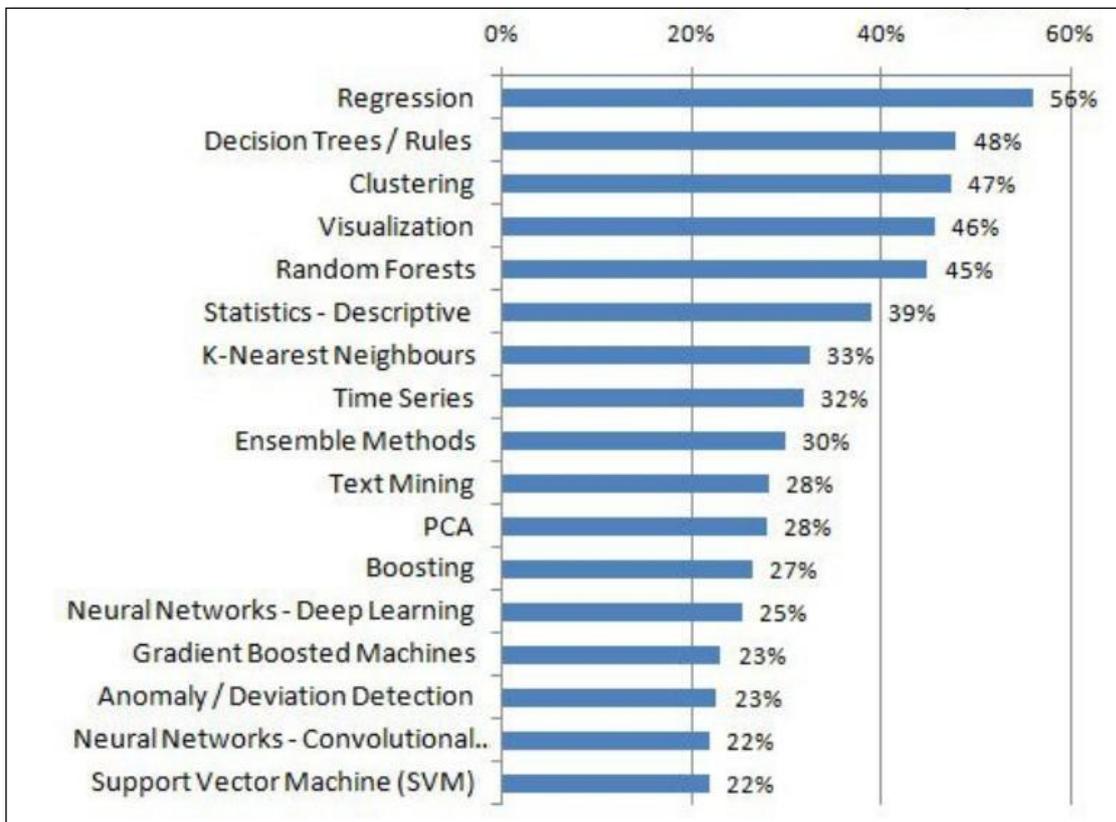


Figura 3 -Técnicas de minería de datos más utilizadas

Fuente: KDnuggets.com

Entre otra de las técnicas se tiene la **Extracción de reglas de asociación**, esta es una metodología que va a permitir comprender la relación entre diferentes cualidades o correlaciones, dentro de una base de datos (Solano, 2016). Este tipo de búsqueda permite discernir y diferenciar mejor a una muestra, elegir qué cualidad o conjunto de cualidades es posible que conlleven a una consecuencia determinada. Posteriormente se encuentran las **Redes bayesianas** que son modelos que representan gráficamente la relación de dependencia probabilística que existe entre diferentes variables. Este tipo de representaciones nos permitirán dar soluciones a diferentes tipos de problemas tanto descriptivos como predictivos.

Luego está la **Regresión**, estos son modelos muy utilizados, que parten de una serie histórica para que a partir de ellas surjan predicciones y formar relaciones entre datos. A través de este método se encuentran anomalías dentro de los datos, permitiendo graficar una línea de evolución trasmisible al futuro.

Finalmente se encuentra el **Modelado estadístico**, este método intenta graficar la relación que existe entre variables explicativas y dependientes, para mostrar los diversos cambios que se muestran a medida que van cambiando ciertos parámetros que se consideración en su formulación, estableciendo una ecuación matemática que pretende reflejar la realidad de una manera muy fiable.

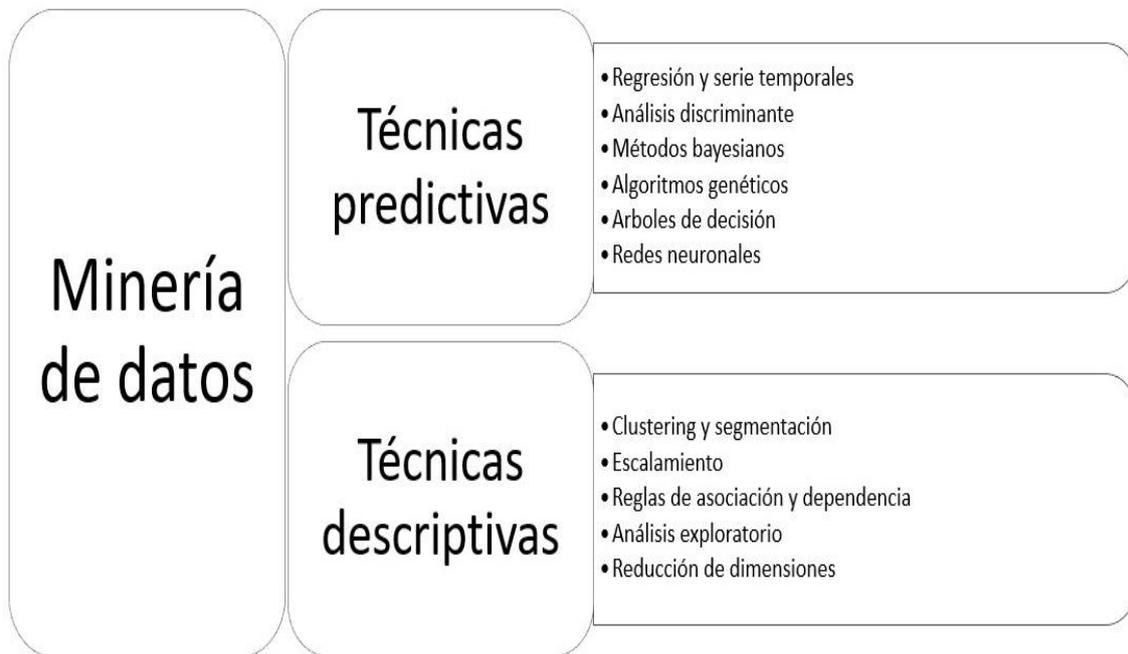


Figura 4 - Técnicas de minería de datos

III. METODOLOGÍA

3.1. Tipo y diseño de investigación:

Tipo de investigación:

El presente trabajo de exploración fue de **tipo aplicada** ya que estuvo basado en la información obtenida en las redes sociales utilizando técnicas de minería de datos que permitieron identificar patrones de consumo de servicios turísticos.

Como nos menciona (Vargas Cordero, 2009) “La investigación de tipo aplicada es aquella que utiliza el conocimiento en la práctica para luego aplicarlos en beneficio de la sociedad en general, todo ello se construye a partir de una necesidad”.

El nivel de profundidad de la presente investigación fue descriptivo, porque se especificó las características, perfiles y comportamientos de un grupo de usuarios. Comprende la explicación, registro, observación e interpretación del medio real y la composición o sucesión de los fenómenos, así como lo refiere (Tamayo y Tamayo, 2003)

La naturaleza de los diversos datos y la información tratada en esta investigación fue de tipo **cuantitativo**; ya que permitió interpretar ciertas características y preferencias de diferentes usuarios en redes sociales con el propósito de descubrir patrones de consumo de servicios turísticos. Como lo señala (Blasco Mira, y otros, 2007) este tipo de investigación estudia la realidad en su ámbito natural y cómo se comporta, observando e interpretando fenómenos de los involucrados.

Por los medios para obtener los datos; esta investigación fue **de laboratorio**, porque se utilizó algunas de las técnicas de minería de datos para obtener los diferentes patrones de consumo de ciertos usuarios en redes sociales. El proceso de investigación se dio dentro de un ámbito equipado, que ayudó al estudio de la materia mediante el uso de las variables y condiciones que influyen. (Rodríguez, 2019).

Diseño de investigación:

Este trabajo exploratorio es de diseño **no experimental**; porque solo se observó ciertos tipos de características y comportamientos que estos usuarios tengan para ciertas preferencias de servicios turísticos. Hernández, Fernández y Baptista (2006) afirman que, en este tipo de investigación, se realiza observando las variables en su realidad tal cual para que luego estas sean analizadas para el estudio.

3.2. Variables y operacionalización:

Variables:

Técnicas de Minería de datos

Definición Conceptual: Proceso mediante el cual se realizó el análisis de grandes bases de datos utilizando diferentes técnicas o herramientas, con la finalidad de establecer patrones, comportamientos, tendencias o características repetitivas, con las cuales se pueda comprender la información obtenida contribuyendo en el proceso de tomar mejores decisiones en un entorno específico.

Subcategorías:

Técnicas de Agrupamiento o Clustering, (similitud o cercanía).

Técnicas de Asociación. (relaciones o dependencias).

3.3. Población, muestra y muestreo:

Con respecto al actual estudio, **la población** de estudio fue La red social Facebook, que tengan una afinidad con los viajes y servicios turísticos.

La muestra en esta investigación se incluyeron usuarios de la red social Facebook que interaccionen con el hashtag #viajesPerú, mediante el cual permitió identificar usuarios quienes tienen una característica en común, la búsqueda de servicios turísticos. Es decir, se incluyeron datos de usuarios con intereses turísticos y se excluyeron los que no tenga esta afinidad.

3.4. Técnicas e instrumentos de recolección de datos:

En este trabajo de exploración se utilizó la **técnica de observación**, ya que se recopiló la información directa de la red social en mención, de usuarios que interaccionen con el hashtag #viajesPerú, esto permitió determinar las expectativas y necesidades que estos usuarios tienen sobre los servicios turísticos. De acuerdo con Rodríguez Peñuelas, Marco (2008) señala que las técnicas son recursos utilizados para recoger información.

El instrumento que se elaboró fue la **guía de observación (ver anexo)**, este es un instrumento de registro que evalúa diferentes características que se establecen como ítems, que permitirán lograr competencias esperadas.

3.5. Procedimientos:

En este trabajo de investigación se procedió a realizar lo siguiente: se elaboró el instrumento para poder realizar la recolección de los datos; que es la guía de observación para registrar ciertos ítems que ayudaron a evaluar ciertas características de los participantes de la muestra de estudio. Se seleccionaron ciertos datos del grupo #viajesPerú que tengan relación con el estudio, donde se observó y analizó ciertos comportamientos de los usuarios pertenecientes a este grupo.

Se analizaron los datos procesados por las técnicas de minería de datos seleccionadas. Se comparará los datos recolectados de cada técnica de minería de datos.

3.6. Método de análisis de datos:

En este trabajo de estudio el método que se utilizó para analizar los datos fue a través de las diversas Técnicas de Minería de datos haciendo uso de R Studio.

3.7. Aspectos éticos:

Por razones éticas como por cuestiones legales, está restringido la publicación de algún tipo de información de los diferentes usuarios de las redes sociales que se ha consultado, así está fuera pública. Todas informaciones de los resultados obtenidos en esta investigación serán dados de forma general, sin revelar información de privacidad de algún usuario.

IV. RESULTADOS:

De acuerdo con los resultados encontrados se tiene que para el primer objetivo:

Identificar patrones de consumo de servicios turísticos en las redes sociales utilizando las técnicas de minería de datos.

Se encontró la información en el grupo #ViajesPerú, en el cual se pudo observar el comportamiento y diferentes reacciones de los usuarios referente a publicaciones, tantos comentarios, compartidos y me gustas; así como también de los meses en que se realizaron las mismas. Dentro de los cuales se tiene: Los usuarios más interesados siempre están preguntando por recomendaciones sobre las publicaciones. Existen en gran parte usuarios que consultan si pueden llevar a niños. Hay publicaciones que tiene más éxito cuando especifican todo sobre sus paquetes e información detallada. Existe una gran tendencia en publicaciones de aventura que tiene buena acogida por los usuarios. Los usuarios interesados en obtener información tanto de los paquetes que ofrecen como lo que incluye y clima de la zona. Existe un número bajo de usuarios que desea incluir a sus mascotas en sus viajes. Existen usuarios que piden paquetes completos.

Se transformaron los datos para que estos puedan ser manejados con las Técnicas de Minería de datos, como se puede observar en la Tabla 1, donde se agrupan los datos según los ítems que se clasificó para las publicaciones correspondientes, los cuales son: Todo incluido, Información y Turismo; de estos se realizó el conteo total de estas reacciones para poder ver el mayor número de reacciones según estos ítems.

MES	TODO INCLUIDO	TURISMO	INFORMACIÓN
ABR	0	0	1659
JUL	0	1622	0
AGOS	5734	2178	0
NOV	822	835	1382

Tabla 1 - Consumo de tipo de publicaciones según Mes

Con lo que corresponde a los objetivos específicos que son:

Establecer las técnicas de minería de datos que se van a utilizar.

Evaluar ciertos patrones para determinar la futura toma de decisiones.

Se procedió a comparar las diversas técnicas de minería de datos más utilizadas, con la finalidad de poder determinar las particularidades entre los datos obtenidos. Posteriormente, luego de haber definido la primera técnica a utilizar, se procedió a desarrollarla, realizando el cálculo del número de Clusters para los agrupamientos correspondientes. Se hizo uso de la herramienta de RStudio, donde se utilizó la técnica de K-means junto con el método del codo para definir el número exacto de clusters a utilizar. Tal como se puede apreciar en la Figura 5, donde se visualiza como número óptimo los 4 primeros grupos; ya que la varianza dentro de cada clúster es mínima y la varianza entre clústeres es mucho más grande.

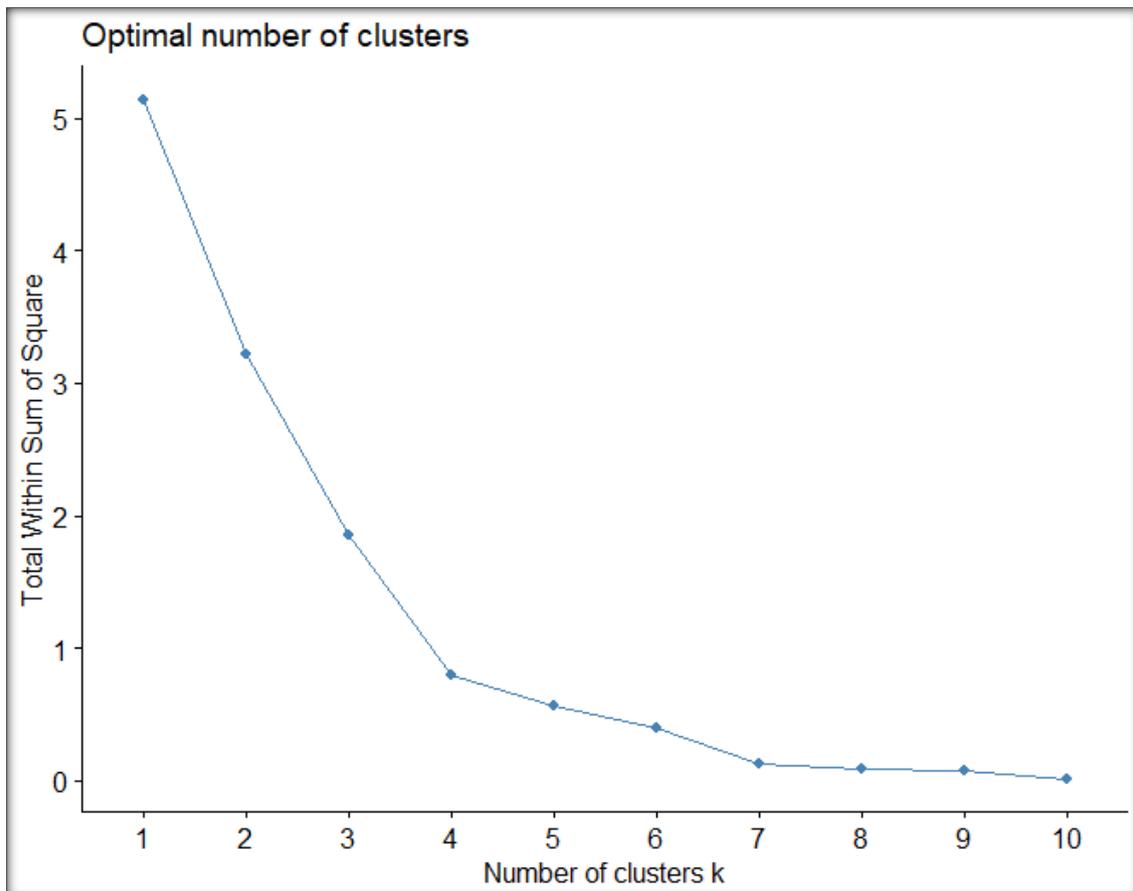


Figura 5 - Cálculo del número óptimo de Clusters

Se obtuvo 4 grupos o clúser (cluster_0, cluster_1, cluster_2 y cluster_3) según la Tabla 2. Grupo de Consumo de Servicio, en la que se puede observar el total de reacciones según el servicio consultado en cada mes.

ABR	JUL	AGOS	NOV	TODO INCLUIDO	TURISMO	INFORMACIÓN	CLUSTERS
1				0	0	1659	cluster_0
	1			0	1622	0	cluster_1
		1		5734	2178	0	cluster_2
			1	822	835	1382	cluster_3

Tabla 2 - Grupo de Consumo según Servicio

Se identificó ciertas características de consumo en cada grupo o clúser clasificado como servicio Todo incluido, Turismo e Información, como se muestra en el Figura 6. Consumo de Servicios según el clúser, en la que se puede observar el tamaño de cada servicio o característica y su distribución de cada clúser.

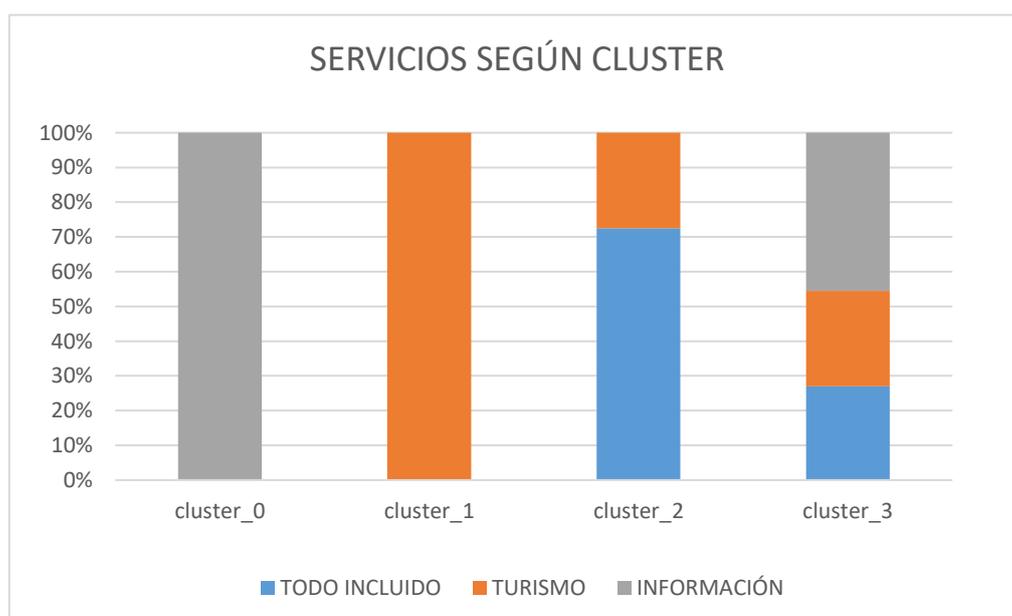


Figura 6 - Grupo de Servicio según Clúster

Se aplicó a los datos recolectados la **Técnica de Minería de Datos de Asociación** y se obtuvo 4 dependencias o relaciones, según Figura 7. Grupo de dependencia de Servicio según la Temporada de la publicación.

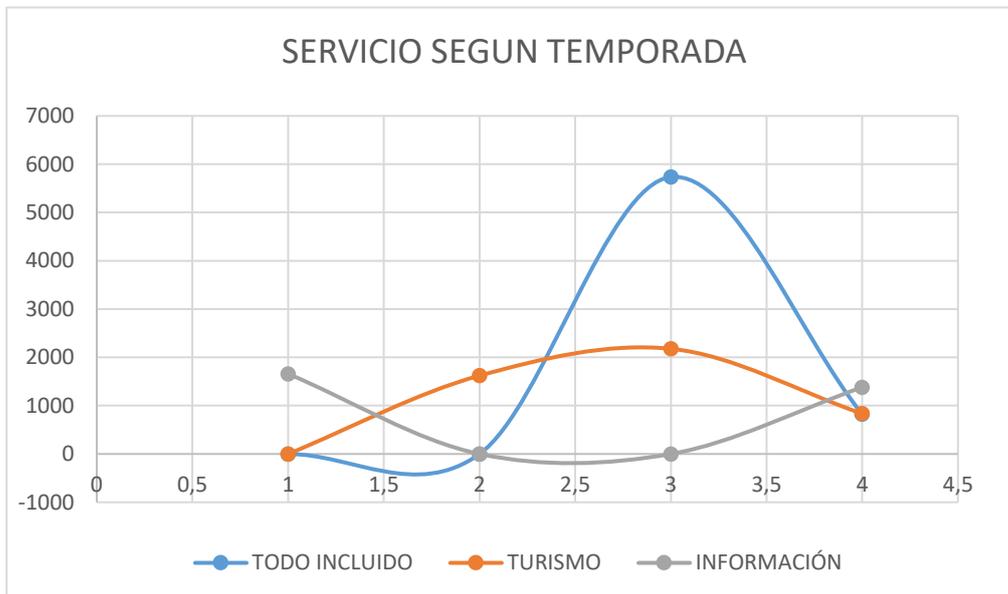


Figura 7 - Grupo de Consumo de Servicios según Temporada

Con la identificación de ciertos patrones y comportamientos que se pusieron a observar y analizar con las siguientes técnicas de minería de datos, se evaluarán los diferentes patrones y con ello determinar futura toma de decisiones que serán una gran ventaja para todas aquellas empresas del sector turismo que estén siempre buscando lo mejor para ofrecer a sus clientes.

V. DISCUSIÓN:

En relación con los hallazgos encontrados en el presente trabajo de tesis; se puede deducir que para el primer objetivo se consiguió identificar ciertos patrones de consumos de paquetes turísticos, haciendo uso de las técnicas de minería de datos; los cuales, permitieron obtener información relevante para conocer más de los usuarios que tienen interés por paquetes turísticos. En este contexto (Fayyad, y otros, 1996) resaltan que la data mining extrae patrones de datos, los cuales son esenciales para generar conocimientos útiles. Así mismo (Leskovec, y otros, 2019) confirman esta hipótesis y exponen que la minería de datos busca modelos o patrones en grandes volúmenes de datos a partir de una distribución subyacente, de la cual se extraen datos visibles. Dentro de lo más resaltante en los resultados detectados, se ha podido determinar que en su gran mayoría los usuarios suelen interesarse por recomendaciones e imágenes que puedan avalar dicho servicio. Lo cual, concuerda con (De Lucca, 2017) que en su investigación "Minería de datos y el turismo", a pesar de plantearse en contexto distinto y hacer uso de diferentes herramientas, afirma que la imagen turística de un destino tiene una gran repercusión por el contenido generado. Con respecto a todo ello (Oliveira, 2013) señala que es de vital importancia conocer y comprender como exactamente los usuarios perciben y manejan la información que se difunde en diferentes medios online; ya que esto permite, desarrollar puntos de atracción en base las exigencias y requisitos detectados. Teniendo en cuenta que muchas de los usuarios o consumidores de paquetes turístico se mueven por el sentir de cada uno; tal como lo mencionan (Hsu CHC, y otros, 2009), donde recalcan que las diferentes tonalidades del turismo varían dependiendo del deseo y la motivación de cada usuario.

Con respecto al segundo objetivo se puede establecer que con la implementación de las técnicas de minería de datos se logró la identificación y clasificación de la información principal para este trabajo de investigación y así poder determinar los patrones de consumo de servicios turísticos. Tal como lo afirman (Arévalo Rodríguez, y otros, 2016) donde en su investigación lograron determinar el patrón de comportamiento de los clientes para el consumo de servicios turísticos, identificando que estos son más frecuentes en determinadas temporadas. Para llegar a estas conclusiones utilizaron técnicas de agrupamiento o Clustering y reglas de asociación, tal como en la presente investigación; pero, además hicieron uso de herramientas adicionales como Google Analytics para extraer información de redes sociales, páginas web y FanPage de las empresas hoteleras de la ciudad. Así mismo, en el estudio de (Córdova Valdivia, y otros, 2018) utilizaron la técnica de árboles de decisión, donde se pudo comprobar que este modelo tiene un gran porcentaje de confiabilidad y asertividad para reducir tiempos y determinar si un alumno será moroso o no. En estas investigaciones se comprobó que las diferentes técnicas de minería de datos permiten alcanzar una pesquisa fundamental teniendo un alto índice de confiabilidad. Respecto a ello (Belinchón Monjas, 2015) sustenta que las técnicas de análisis de datos con data mining ponen al alcance el conocimiento que se necesita para que las actividades sean mucho más efectivas. Así mismo (Mahadev Ramageri, 2010) refuerza esta teoría mencionando que las organizaciones hoy en día están haciendo uso de las diversas técnicas de data mining para descubrir conocimiento y mejorar sus negocios, por lo que vienen encontrando excelentes resultados.

Por otra parte, teniendo en cuenta la información proporcionada a partir de la identificación de patrones de consumo de servicio turístico detectados usando técnicas de minería de datos; para el tercer objetivo, se pudo inferir que dichas referencias permiten tener conocimiento a cerca de las características y exigencias de los consumidores de servicios, para tomar decisiones y de esta manera poder generar valor agregado para crear ventajas competitivas. Tal como indica (Uriarte del Águila, 2018) en su investigación que, haciendo uso de la técnica de reglas difusas, se alcanzó a reducir la gestión de toma de decisiones a solo 2, disminuyendo el tiempo para el mismo; aseverando así que la minería de datos permite obtener información relevante para ayudar o mejorar la toma de decisiones. Esto permitió concluir que, a pesar de hacer uso de técnicas diferentes en comparación al presente estudio, se puede lograr el mismo objetivo; facilitar pesquisa importante para la toma de decisiones. Tal como lo afirman (Talavera Pereira, y otros, 2007) donde infieren que el valor real de los datos radica cuando estos se analizan y contribuyen a la toma de decisiones y, sobre todo, ayudan a mejorar el entendimiento de los fenómenos que nos rodean. También, indican que la minería de datos es la sucesión completa que permite extraer información, para poder explorarla de manera directa y luego realizar el análisis respectivo para convertirla en conocimiento y de esta manera ser soporte en la toma de decisiones. (García, y otros, 2006) concuerdan con esta definición y señalan que la minería de datos a través de la búsqueda y cuantificación de relaciones predictivas y descriptivas a partir de datos brutos permite transformar los datos en conocimiento para el uso en la toma de decisiones. Por su lado (Salinas Flores, 2000) permite sustentar estas afirmaciones ya que, asegura que las diferentes herramientas de minería de datos pronostican los nuevos aspectos y predicen situaciones para ayudar a la toma de decisiones de negocio proactivamente.

VI. CONCLUSIONES

En relación con los resultados obtenidos durante el desarrollo de la presente investigación se llegó a las siguientes conclusiones:

1. Se concluye que se encontraron patrones de consumo de paquetes turísticos de los diferentes usuarios que interactúan en la red social Facebook con el hastag #viajesPerú. Tales patrones se identificaron de un conjunto de datos, los cuales, para poder ser manejados con las diferentes técnicas de minería de datos, se manipularon mediante un proceso de selección y tabulación que permitieron llegar agruparlos según cada ítem de los cuales se clasificó las publicaciones identificadas.
2. Se concluye también que las técnicas de minería de datos seleccionadas permitieron comprobar que se puede utilizar diferentes tipos de filtros. Dado que con la técnica de Clustering se logró determinar ciertos grupos de datos; que debido a la similitud o dependencia que en estos existe, ayudó a determinar los servicios que los usuarios consumen y así mismo con la aplicación de la Técnica de Asociación permitió obtener grupos de dependencias o relaciones, según la Temporada de cada una de las publicaciones.
3. Se concluye además que en esta investigación queda en evidencia una vez más la importancia y relevancia de los beneficios que proporciona la minería de datos, ya que hoy en día la data mining permite procesar no solo datos cuantitativos, sino que también permite evaluar y transformar datos cualitativos en datos numéricos. Los cuales permiten que las empresas que se dedican no solo al rubro del turismo, sino que en cualquier ámbito; puedan mejorar sus estrategias de negocio y sobre todo contar con herramientas que les facilite la toma de decisiones para generar mejoras continuas dentro del entorno en el que se desempeñan.

VII. RECOMENDACIONES

1. Se recomienda para futuras investigaciones hacer uso de minería de datos en otros ámbitos sociales como la salud y bienestar, donde se puede aplicar modelos predictivos, para el diagnóstico de diversas enfermedades, sirviendo como apoyo para los especialistas de las diferentes ramas de la medicina.
2. Se sugiere para futuros estudios, el uso de otras técnicas de minería de datos como redes bayesianas, así como también de redes neuronales, las cuales son poderosas herramientas de aprendizaje automático que permiten analizar datos en grandes volúmenes y así mismo permiten resolver problemas tanto descriptivos como predictivos.
3. Se recomienda también a futuros investigadores hacer uso de otras herramientas como Google Analytics para el proceso inicial de los datos, ya que esta herramienta permite y garantiza una mejor comprensión de dicha data inicial.
4. Se sugiere utilizar otras plataformas digitales o fuentes de datos para el recojo de información, así como también el uso de # (hashtag) o filtros relacionados con otros temas como la prevención y protección de menores, violencia domestica o cualquier otro tipo de violencia en general, los cuales facilitarían información fundamental a las autoridades pertinentes para las medidas del caso. Toda esta data es posible haciendo uso de la minería de texto, ya que se puede analizar a través de las palabras o emojis el sentir o el estado de ánimo de las personas que interactúen en las diversas plataformas, donde se puede tener los indicios para detectar dichas casuísticas.

5. Finalmente se recomienda hacer uso de diferentes herramientas que puedan permitir un mejor entendimiento de los datos o realizar la obtención de información de forma más rápida. Tal como lo es R Studio, una herramienta Open Source, que facilita a través de sus diversas librerías la extracción de grandes cantidades de datos de diferentes plataformas. Así mismo se recomienda el uso de Pentaho que es una herramienta potente de BI que mejora la capacidad para analizar datos, pero que sobre todo mejora la capacidad de toma de decisiones debido a la flexibilidad que este maneja.

REFERENCIAS BIBLIOGRÁFICAS

El Peruano. 2022. Diario el Peruano . [En línea] 2022. <https://elperuano.pe/noticia/138297-sector-turismo-aportaria-25-al-pbi-el-presente-ano>.

Han, Jiawei, Pei, Jian y Kamber, Micheline . 2011. *Data Mining: Concepts and Techniques*. 2011. ISBN 978-0-12-381479-1.

Aggarwal, Charu C. 2018. *Social Network Data Analytics*. New York : s.n., 2018. e-ISBN 978-1-4419-8462-3.

Alania Ricaldi, Pit Frank . 2018. Aplicación de Técnicas de Minería de Datos para Predecir la deserción estudiantil de la Facultad de Ingeniería de la Universidad Nacional Daniel Alcides Carrión. Pasco, Perú. [En línea] 2018. http://repositorio.undac.edu.pe/bitstream/undac/829/1/T026_40573846_M.pdf.

Alhadj , Reda y Rokne, Jon . 2018. *Encyclopedia of Social Network Analysis*. [ed.] Springer Science+Business Media. 2018. ISBN 978-1-4939-7131-2 (eBook).

Antonietti, Betiana. 2017. Servicios Turísticos I. [En línea] 2017. <http://www.utntyh.com/wp-content/uploads/2016/09/OIT.pdf>.

Arévalo Rodríguez, Tania y Noriega Coronel, Mario . 2016. Minería de Datos para la Identificación de patrones de consumo de servicios turísticos para Orientar la oferta en el sector hotelero de la ciudad de Trujillo. *Repositorio institucional de la Universidad Privada Antenor Orrego*. [En línea] 2016. <https://repositorio.upao.edu.pe/handle/20.500.12759/5739> .

Aušrinė. 2021. Social Media Data Mining: What It Is, How It Works, and How to Use It. [En línea] 2021. <https://whatagraph.com/blog/articles/social-media-data-mining>.

Begoña Munarriz. 1992. Técnicas y métodos en Investigación cualitativa. *Universidad de Vasco*. 1992.

Belinchón Monjas, Yolanda . 2015. Minería de datos. Universidad Carlos III de Madrid. 2015.

Blasco Mira, J y Pérez Turpin, J. 2007. Naturaleza y enfoques de la investigación cualitativa. En Metodologías de Investigación en las Ciencias de la Actividad Física y el Deporte: Ampliando Horizontes (309). San Vicente, Alicante, España : Editorial Club Universitario, 2007.

Bucaram, Santiago, Jorge , Daries y Jaime, Vicente. 2019. Evolución del turismo en Perú 2010 - 2020, la influencia del COVID -19 y recomendaciones pos -COVID -19. [En línea] 2019. <https://publications.iadb.org/publications/spanish/document/Evolucion-del->

turismo-en-Peru-2010-2020-la-influencia-del-COVID-19-y-recomendaciones-pos-COVID-19-nota-sectorial-de-turismo.pdf.

Córdova Valdivia, Alenster y Torres Jurado, Karen . 2018. Aplicación de la Minería de Datos para pronosticar el Riesgo de Morosidad de los estudiantes de la Universidad Autónoma del Perú. . Lima, Perú : s.n., 2018.

Cuñat Giménez, Rubén. 2007. Aplicación de la Teoría fundamentada (Grounded Theory) al estudio del Proceso de Creación de Empresas. 11.11.21, de Dialnet. 2007.

De Lucca, Magdalena. 2017. Data Mining y turismo: Un enfoque integral para potenciar la imagen de Argentina. *Repositorio digital de la Universidad Nacional del Sur.* [En línea] 2017. <https://repositoriodigital.uns.edu.ar/bitstream/handle/123456789/4035/Tesis%20Magdalena%20De%20Lucca.pdf?sequence=1&isAllowed=y>.

Donald Jense, David y Neville , Jade . 2002. Conference: Invited presentation to the National Academy of Sciences Workshop on Dynamic Social Network Modeling and Analysis. Data Mining in Social Networks. [En línea] 2002. https://www.researchgate.net/publication/228057937_Data_mining_in_social_networks.

Fayyad, Usama , Piatetsky-Shapiro, Gregory y Smyth, Padhraic . 1996. From Data Mining to Knowledge Discovery in Databases. s.l. : AI Magazine, 1996. Vol. 17, 3.

Figuroa Pinedo, Jessica Ruth. 2019. Développement touristique ou reproduction sociale de la pauvreté. [En línea] 2019. <https://www.erudit.org/fr/revues/teoros/2014-v33-n2-teoros03306/1042431ar/>.

Figuroa Pinedo, Jessica Ruth. 2018. Gestión turística y desarrollo sostenible en sitios Patrimonio de la Humanidad. Una mirada al caso de Machu Picchu (Cuzco-Perú). *Dialnet.* [En línea] 2018. <https://dialnet.unirioja.es/servlet/articulo?codigo=6640362>.

Fuller, Norma. 2009. *Turismo y cultura. Entre el entusiasmo y el recelo.*. Lima : Fondo Editorial Pontificia Universidad Católica del Perú, 2009. Registro de proyecto editorial N: 31501360900364.

Garcia, Juan Carlos y Zanfrillo , Alicia Inés . 2006. Data mining application to decision-making processes in University Management. Mar de Plata, Argentina : s.n., 2006.

Guerrero González, Perla Elizabeth y Ramos Mendoza, José Roberto. 2014. *Introducción al Turismo.* México : IBSN e-book: 978-607-438-875-6, 2014.

Hamilton, Laura. 2014. Six Novel Machine Learning Applications. [En línea] 2014. <https://www.forbes.com/sites/85broads/2014/01/06/six-novel-machine-learning-applications/?sh=5fc0482f1060>.

Hsu CHC, Cathy, A. Cai, Liping y Li, Mimí . 2009. Expectativa, motivación y actitud: un modelo de comportamiento turístico. *Revista de investigación de viajes*. 2009.

Insightsoftware. 2019. ¿Cuál es la función de la minería de datos en la business intelligence? *Insightsoftware Spain [en línea]*. [En línea] 2019. <https://insightsoftware.com/es/blog/what-role-does-data-mining-play-for-business-intelligence/>.

Jaramillo Valbuena, Sonia, Cardona, Sergio Augusto y Fernández, Alejandro . 2015. *Minería de datos sobre streams de redes sociales, una herramienta al servicio de la Bibliotecología. Información, cultura y sociedad*. [En línea] 2015. <https://www.redalyc.org/articulo.oa?id=263042678005>.

KDnuggets. 2019. KDnuggets. *Top Data Science and Machine Learning Methods Used in 2018, 2019*. 2019.

Leskovec, Jure , Rajaraman, Anand y D. Ullman, Jeffrey . 2019. Mining of Massive Datasets. [En línea] 2019. <http://www.mmids.org/>.

Loy, James . 2019. *Neural Network Projects with Python*. [ed.] James Loy. Birmingham : Packt Publishing Ltd., 2019. ISBN 978-1-78913-890-0.

Mahadev Ramageri, Bharati . 2010. DATA MINING TECHNIQUES AND APPLICATIONS. *Indian Journal of Computer Science and Engineering*. 2010. Vol. 1, 4, págs. 301-305.

Microsoft. 2022. Conceptos de minería de datos. [En línea] 2022. <http://msdn.microsoft.com/es-es/library/ms174949.aspx>.

MINCETUR. 2018. Ministerio de Comercio Exterior y Turismo . [En línea] 2018. <https://www.mincetur.gob.pe/wp-content/uploads/documentos/turismo/tripticos/2018/Piura.pdf>.

Oladipupo, Taiwo. 2014. Types of Machine Learning Algorithms. [En línea] 2014. <http://cdn.intechopen.com/pdfs-wm/10694.pdf>.

Oliveira, Eduardo. 2013. The digital complexity in destination branding: the case of Portugal as tourism destination. [En línea] 2013. https://mpra.ub.uni-muenchen.de/52680/1/MPRA_paper_52680.pdf.

OMT. 2020. Organización Mundial del Turismo . [En línea] 2020. <https://www.unwto.org/es/el-turismo-mundial-consolida-su-crecimiento-en-2019>.

Palma, Claudio, Palma, Wilfredo y Pérez, Ricardo . 2009. *Data Mining el arte de anticipar*. Santiago de Chile : Ril editores, 2009. ISBN: 978-956-284-711-7.

PROMPERÚ. 2016. Comisión de Promoción del Perú para la Exportación y el Turismo. *Evaluación del Turismo Rural Comunitario en el Perú*. [En línea] 2016. https://www.promperu.gob.pe/TurismoIN/Uploads/mercados_y_segmentos/segmentos/1031/TRCPeru2016.pdf.

Rivas, Esther. 2018. ¿Qué es el Data Mining o minería de datos? [En línea] 2018. <https://www.iebschool.com/blog/data-mining-mineria-datos-big-data/>.

Rodríguez Peñuelas, Marco Antonio. 2010. Métodos de Investigación. Universidad Autónoma de Sinaloa. México. 2010.

Rodríguez, Daniela. 2019. Investigación de laboratorio: qué estudia, tipos, ventajas, desventajas. Liferder. 26 de julio de 2019.

Salinas Flores, Jesús . 2000. Toma de Decisiones con Data Mining. *Industrial Data*. Lima, Perú : s.n., 2000. Vol. 3, 2, págs. 80-82.

Solano, Braulio. 2016. Tareas de la minería de datos: reglas de asociación y secuencias. [En línea] 2016. <http://bsolano.com/ecci/claroline/backends/download.php/UHJlc2VudGFjaW9uZXMvNy5fVGFyZWZzX2RlX2xhX21pbmVy7WFfZGVfZGF0b3MsX3JlZ2xhc19kZV9hc29jaWFjafNuLnBkZg%3D%3D?cidReset=true&cidReq=CI2352>.

Statista. 2021. Statista GmbH. [En línea] 2021. <https://es.statista.com/estadisticas/640133/aportacion-del-sector-turistico-al-pib-mundial/>.

Talavera Pereira, Rosalba y Arcano Aular, Yelitza Josefina. 2007. Minería de Datos como soporte a la toma de decisiones empresariales. Maracaibo, Venezuela : s.n., 2007. Vol. 23, 52, págs. 104-118. ISSN: 1012-1587..

Tamayo y Tamayo, Mario . 2003. El proceso de la investigación científica. [En línea] 2003. https://www.academia.edu/17470765/EL_PROCESO_DE_INVESTIGACION_CIENTIFICA_MARIO_TAMAYO_Y_TAMAYO_1.

Tech-Experts. 2014. Difference Between Data Mining and Machine Learning. [En línea] 20 de 05 de 2014. <http://dash10mesh.blogspot.com/2014/05/difference-between-data-mining-and.html>.

The World Bank Group. 2017. *Tourism for Development*. Washington, DC : s.n., 2017.

Uriarte del Águila, Chanith . 2018. Minería de datos para la toma de decisiones en el área de gestión al cliente de telefónica del. [Tesis para optar el título de Ingeniero de Sistemas e Informática.] Repositorio institucional de la Universidad Nacional de San Martín - Tarapoto. [En línea] 2018. <http://repositorio.unsm.edu.pe/handle/11458/2682>.

Vargas Cordero, Zoila Rosa. 2009. La Investigación Aplicada: Una forma de conocer las realidades con evidencia científica. 2009.

Viera, Luis, Ortiz, Luis y Ramirez, Santiago. 2009. *Introducción a la minería de datos* . Rio de Janeiro : E-papers Servicios Editoriales, 2009.

Witten, Ian H, Frank, Eibe y Hall, Mark A. 2011. *Data mining: practical machine learning tools and techniques* . s.l. : IBSN: 978-0-12-374856-0, 2011.

ANEXOS:

1. Matriz de Categorización:

CATEGORÍAS	DEFINICIÓN CONCEPTUAL	SUB CATEGORÍAS	RASGOS
Técnicas de Minería de Datos	Es el conjunto de técnicas y tecnologías que permiten explorar grandes bases de datos (Tech-Experts, 2014), de manera automática o semiautomática, con el objetivo de encontrar patrones repetitivos, tendencias o reglas que expliquen el comportamiento de los datos en un determinado contexto.	Técnicas de Clustering o Agrupamiento	Cercanía o similitud
		Técnicas de Asociación	Dependencia Relaciones

2. Guía de Observación: Registro de incidencia dentro del grupo de usuarios de la red social:

#ViajesPerú – Facebook

GUIA DE OBSERVACIÓN				ITEM
1. ¿Sin planes para Año Nuevo 2022? 12 espacios disponibles – Tour: Oxapampa + Pozuzo + Valle del Perené (21-11)	17	10	140	Todo incluido
2. Vamos a visitar el Nevado Rajuntay y sus hermosos paisajes (04-08)	530	1108	61	Turismo
3. Enamórate de Piura: Máncora – Zorritos – Tumbes (17-11)	203	25	189	Todo incluido
4. Nevado Rajuntay + Trekking (06-11)	53	26	159	Todo incluido
5. Club Campestre Agualandia – Carabayllo – Lima (27-08)	2108	3047	579	Todo incluido
6. Volvimos con el 2x1 – Destinos imperdibles de Cusco (05-11)	59	31	201	Turismo
7. Full Day – Nevado de Rajuntay (11-08)	119	44	316	Turismo
8. Canta Cordillera de la Viuda (13-07)	321	96	1205	Turismo
9. Las bellas aguas termales de Cocalmayo son una de las mejores pozas termales en la región Cusco. (13-11)	30	233	637	Información
10. Marcahuasi. (22-11)	14	0	95	Información
11. Seguro que han oído hablar de los baños termales de Guñoc (18-11)	50	8	315	Información
12. Club campestre Las Gambusinas. (13-11)	242	141	161	Turismo
13. Hacienda Pampalimena – Distrito de Palcazú (26-04)	141	400	1118	Información
 = COMENTARIOS  = COMPARTIDOS  = ME GUSTAS				

3. Certificado de Traducción de Abstract

This document has been translated by the Translation and Interpreting Service of Cesar Vallejo University and it has been revised by the native speaker of English: Mark Stables.



A handwritten signature in black ink, reading "Ana Gonzales Castañeda".

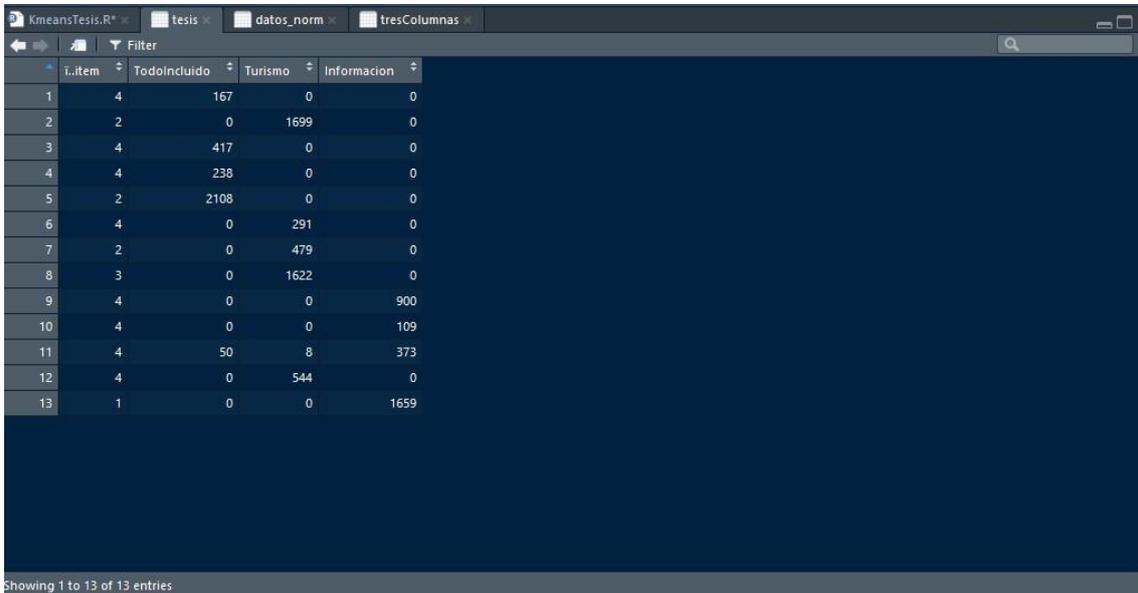
Dr. Ana Gonzales Castañeda
Professor of the School of Languages

4. Tratamiento de datos en RStudio

- Importación de librerías y del archivo con los datos recolectados para el tratamiento de datos en RStudio

```
KmeansTesis.R* | tesis | datos_norm | tresColumnas |
Source on Save | Run | Source
1 library(readr)
2 library(factoextra)
3 library(ggplot2)
4 library(NbClust)
5 library(tidyr)
6
7 tesis = read.table("tesis.csv", header = TRUE , sep = ";")
8 View(tesis)
9
10 tesis$item = as.factor(tesis$item)
11 tesis$item = unclass(tesis$item)
12
13
14 #tesis$Mes = as.factor(tesis$Mes)
15 #tesis$Mes = unclass(tesis$Mes)
16
17
18 #tesis$item = as.factor(tesis$item)
19 #tesis$item = unclass(tesis$item)
20 min_max_norm <- function(x){
21   (x - min(x))/(max(x)-min(x))
22 }
23
24 #convertimos datos a numericos
25 #tesis$item.Nro = as.numeric(tesis$item.Nro)
26
27 #tesis$item = as.numeric(tesis$item)
28 #tesis$item = as.numeric(tesis$item)
29 #tesis$comentario = as.numeric(tesis$comentario)
30 #tesis$compartido = as.numeric(tesis$compartido)
31
```

- Visualización de los datos



The screenshot shows the RStudio interface with a data table displayed in the Environment pane. The table has four columns: 'i.item', 'TodoIncluido', 'Turismo', and 'Informacion'. The data is as follows:

i.item	TodoIncluido	Turismo	Informacion
1	4	167	0
2	2	0	1699
3	4	417	0
4	4	238	0
5	2	2108	0
6	4	0	291
7	2	0	479
8	3	0	1622
9	4	0	900
10	4	0	109
11	4	50	8
12	4	0	544
13	1	0	0

Showing 1 to 13 of 13 entries

- Proceso de normalización de datos para su posterior análisis con las técnicas de minería de datos.

```

33
34
35 #convertimos datos a numericos
36 #tesis$i..Mes = as.numeric(tesis$i..Mes)
37 tesis$TodoIncluido = as.numeric(tesis$TodoIncluido)
38 tesis$Turismo = as.numeric(tesis$Turismo)
39 tesis$Informacion = as.numeric(tesis$Informacion)
40
41
42 # datos normalizados
43 datos_norm = as.data.frame(lapply(tesis, min_max_norm))
44 View(datos_norm)
45

```

- Datos normalizados.

i.item	TodoIncluido	Turismo	Informacion
1	1.0000000	0.07922201	0.00000000
2	0.3333333	0.00000000	1.00000000
3	1.0000000	0.19781784	0.00000000
4	1.0000000	0.11290323	0.00000000
5	0.3333333	1.00000000	0.00000000
6	1.0000000	0.00000000	0.17127222
7	0.3333333	0.00000000	0.281930547
8	0.6666667	0.00000000	0.954679223
9	1.0000000	0.00000000	0.00000000
10	1.0000000	0.00000000	0.00000000
11	1.0000000	0.02371917	0.004708652
12	1.0000000	0.00000000	0.320188346
13	0.0000000	0.00000000	0.00000000

- Se procede a generar una matriz de distancia para luego poder calcular en número óptimo de clusters (Figura 5).

```

45
46 # creamos matriz de distancia
47
48
49 matriz_distancia = get_dist(datos_norm, method = "euclidian")
50 fviz_dist(matriz_distancia, gradient = list(low="white", high="blue"))
51
52 # estimamos clusters
53 fviz_nbclust(datos_norm, kmeans, method = "wss")
54
55
56
57

```

- Luego se estiman las medias de los clusters usando las variables originales.

```

54
55
56
57
58 numeroCluster = NbClust(datos_norm , distance = "euclidean", min.nc = 2 , max.nc = 6, method = "kmeans", index = "allc
59
60 dosclusters = kmeans(datos_norm, centers = 3,nstart= 100)
61 dosclusters
62
63 cuatroclusters = kmeans(datos_norm, centers = 4,nstart= 100)
64 cuatroclusters
65
66 fviz_cluster(cuatroclusters,data = datos_norm)
67
68 # estimamos las medias de los clusters usando las variables originales
69
70 tesis %>% mutate(cluster = cuatroclusters$cluster) %>% group_by(cluster) %>% summarise_all("mean")
71 datos_norm$cluster = as.factor(cuatroclusters$cluster)
72
73 trescolumnas =gather(datos_norm, Atributo, valor, TodoIncluido:Informacion, factor_key = TRUE)
74
75 View(trescolumnas)
76
77 ggplot(trescolumnas, aes( as.factor(x= Atributo),
78                          y=valor,
79                          group=cluster, colour= cluster))+
80   stat_summary(fun = mean, geom = "pointrange", size = 1)+
81   stat_summary(geom = "line")
82
83

```

- Se procede agrupar los datos en tres columnas para tener todas las combinaciones posibles por cada cluster.

i.item	cluster	Atributo	valor
1	1.0000000	3	0.079222011
2	0.3333333	1	0.000000000
3	1.0000000	3	0.197817837
4	1.0000000	3	0.112903226
5	0.3333333	2	1.000000000
6	1.0000000	3	0.000000000

i.item	cluster	Atributo	valor
7	0.3333333	1	TodoIncluido 0.000000000
8	0.6666667	1	TodoIncluido 0.000000000
9	1.0000000	3	TodoIncluido 0.000000000
10	1.0000000	3	TodoIncluido 0.000000000
11	1.0000000	3	TodoIncluido 0.023719165
12	1.0000000	3	TodoIncluido 0.000000000
13	0.0000000	4	TodoIncluido 0.000000000
14	1.0000000	3	Turismo 0.000000000
15	0.3333333	1	Turismo 1.000000000
16	1.0000000	3	Turismo 0.000000000
17	1.0000000	3	Turismo 0.000000000
18	0.3333333	2	Turismo 0.000000000
19	1.0000000	3	Turismo 0.171277222
20	0.3333333	1	Turismo 0.281930547
21	0.6666667	1	Turismo 0.954679223

i.item	cluster	Atributo	valor
22	1.0000000	3	Turismo 0.000000000
23	1.0000000	3	Turismo 0.000000000
24	1.0000000	3	Turismo 0.004708652
25	1.0000000	3	Turismo 0.320188346
26	0.0000000	4	Turismo 0.000000000
27	1.0000000	3	Informacion 0.000000000
28	0.3333333	1	Informacion 0.000000000
29	1.0000000	3	Informacion 0.000000000
30	1.0000000	3	Informacion 0.000000000
31	0.3333333	2	Informacion 0.000000000
32	1.0000000	3	Informacion 0.000000000
33	0.3333333	1	Informacion 0.000000000
34	0.6666667	1	Informacion 0.000000000
35	1.0000000	3	Informacion 0.542495479
36	1.0000000	3	Informacion 0.065702230

i..item	cluster	Atributo	valor
37	1.0000000	3	Informacion
38	1.0000000	3	Informacion
39	0.0000000	4	Informacion

- Finalmente se observa el comportamiento según la dependencia de los grupos tratados donde se visualiza las 4 dependencias encontradas para los servicios de Todo incluido, servicio de turismo y el servicio de información.

