Facultad de Ingeniería

Trabajo de Investigación

“Aprendizaje automático para la optimización de procesos de marketing digital en el sector turístico”

Autores:

Flores Limaylla, Guadalupe Isabel 1611616
Peña Alvarez, Eddy Paolo U17102502

Para obtener el Grado Académico de Bachiller en:

Ingeniería de Software

Lima, Julio 2020

El presente trabajo de investigación está dedicado a las personas que más han influenciado en nuestras vidas, apoyándonos y guiándonos en el transcurso del desarrollo de nuestra profesión.

ÍNDICE DE CONTENIDO

ÍNDICE DE TABLAS

ÍNDICE DE IMÁGENES

CAPITULO I: RESUMEN

1.1. TÍTULO

Aprendizaje automático para la optimización de procesos de marketing digital en el sector turístico

1.2. RESUMEN

El presente trabajo de investigación tiene el objetivo de realizar un estudio de los modelos utilizados en las metodologías de machine learning para la optimización de procesos, enfocándonos en específico en el sector turismo por ser uno de los ejes más importantes en la economía mundial.

Para el desarrollo de la investigación se ha realizado estudios de distintos artículos de investigación que se encuentran relacionados a los temas relacionados a este trabajo, con la finalidad de argumentar la evaluación de los diferentes tipos de machine learning en base a los resultados que se obtuvieron en estudios realizados por investigadores, para conocer el rendimiento y la precisión en referencia a la clasificación de opiniones

de usuarios publicados en páginas web de turismo.

Palabras Clave: machine learning, inteligencia artificial, análisis de sentimiento, turismo, marketing y minería de texto.

CAPITULO II: INTRODUCCIÓN

2.1. CONTEXTO

En las últimas décadas la evolución de la tecnología ha permitido que diversos dispositivos de información se desarrollen, tales como teléfonos móviles y computadores personales, los cuales actualmente se encuentran al alcance de todos. Según lo mencionado por Rueda (2007) “la red ya se puede ver en tiempo real el sentir de la humanidad, pero al mismo tiempo también es posible tergiversar, manipular o frivolizar este sentir, es decir, paradójicamente, los medios de comunicación también pueden usarse para separar y aislar” (p. 12). Ello quiere decir, que se ha generado una dependencia tecnológica con los medios de información abierta que se encuentran publicados en la red, en donde se pueden consultar desde información confiable, hasta en algunos casos información falsa.

En esta nueva era, se han producido muchos cambios en las formas de cubrir las necesidades, donde en su mayoría las personas se han adaptado con facilidad el estar conectados a la red. Con ello, hacemos referencia al sin fin de actividades

(recibir cursos online, realizar compras por internet, hacer pagos, transferencias bancarias, etc.) que podemos realizar desde un computador, una Tablet y hasta un teléfono celular. En los últimos 10 años, según ICEMED (Instituto de la Economía Digital de ESIC), la mayoría de las empresas, sin importar el rubro, consideran realizar publicidad mediante la red como una forma de abordar la aproximación con sus usuarios (PuroMarketing, 2019). Todo ello indica una guerra por conseguir grandes cantidades de clientes; lo que conlleva a las empresas a plantearse diversos métodos para capturar la atención de sus clientes mediante procesos de marketing.

En el Perú, según el Ministerio de Comercio Exterior y Turismo – MINCETUR (2018), la llegada de turistas internacionales ese año alcanzó los 4,4 millones generando 4,895 millones de dólares en ingresos de divisas. Asimismo, se registró 55,4 millones en arribos a los establecimientos de hospedaje.

Debido a la creciente demanda de servicios en el sector, se genera grandes cantidades de información por lo cual las empresas del sector turismo requieren de herramientas y metodologías tecnológicas con la finalidad de automatizar sus procesos para la clasificación de datos lo cual servirá para optimizar la toma de decisiones. Considerándose las opiniones de los usuarios el activo más importante en el empleo de estrategias comerciales, que permitan mejor su aceptación en el mercado.

Los desarrollos de las nuevas tecnologías para gestionar la información digital en el área de marketing en empresas del sector turístico han permitido que los medios tradicionales hayan sido desplazados (Baggio, D'Amico, Llena, Puorto y Travaglini, 2016).

2.2. OBJETIVOS DE LA INVESTIGACIÓN

Los objetivos y el alcance que sustentan el desarrollo del trabajo de investigación, y permitirán responder la pregunta planteada anteriormente, son los siguientes:

2.2.1. OBJETIVO GENERAL

Analizar los modelos de aprendizaje automático para optimizar el proceso de marketing digital en el sector turístico.

2.2.2. OBJETIVOS ESPECÍFICOS

- Revisar diferentes fuentes bibliográficas relacionadas a la implementación de aprendizaje automático basados en comentarios de consumidores dentro del sector turismo.
- Comparar los resultados obtenidos de las técnicas de clasificación.
- Identificar el modelo de aprendizaje automático para la optimización del proceso de marketing digital.

2.3. PREGUNTA DE INVESTIGACIÓN

La presente investigación se desarrolla para el área de marketing en las empresas del sector turístico, debido a su importancia dentro de las organizaciones. Razón por la cual, se busca responder la siguiente pregunta planteada para el desarrollo de la investigación:

- ¿Qué modelos de aprendizaje automático optimizan el proceso de marketing digital en el sector turismo?

2.4. ESTRUCTURA DEL TEXTO

Con el motivo de estructurar de forma ordenada el desarrollo del proyecto se dividió en 6 capítulos el presente trabajo de investigación, los cuales se organizan según el índice mostrado en la página 4 del documento.

CAPÍTULO III: METODOLOGÍA

3.1. FUENTES DE INFORMACIÓN

En los siguientes párrafos se mencionarán las diferentes fuentes de información que se emplearán en el desarrollo de la presente investigación, para la búsqueda de estudios relacionados a la implementación de modelos de machine learning en los procesos de marketing en el sector turismo, encontrados en diferentes repositorios (Ver tabla 1).

Tabla 1: Repositorios de artículos de investigación

REPOSITORIO	DESCRIPCIÓN
Scopus	Es un repositorio de resúmenes y citas literarias extraídas de libros y revistas científicos y actas de congreso.
IEEE XPLORE	Fuente de datos de investigaciones académicas que brinda facilidad de acceso a trabajos de investigación y artículos sobre ciencia.
Scielo	Biblioteca electrónica que contiene colecciones de revistas y artículos científicos.
Journal of Travel Research	Revista de investigación orientada en el comportamiento, desarrollo y gestión de viajes y

	turismo.
Google Scholar	Es un buscador para documentos con orientaciones académicas de diversos libros, artículos y tesis.

Las diversas investigaciones seleccionadas emplean modelos de machine learning distintos para el análisis de opiniones públicas en la web relacionados al sector turismo con respecto a hoteles, restaurantes y lugares turísticos. Ello, nos permitirá realizar un comparativa entre modelos según el rendimiento de las técnicas y algoritmos de machine learning, optando por la selección de los resultados más óptimos y con altos índices de exactitud para la clasificación de opiniones según sus sentimientos.

3.2. CRITERIOS DE BÚSQUEDA

Se utilizó para la búsqueda de los artículos de investigación, palabras claves, tales como: machine learning, inteligencia artificial, análisis de sentimiento, turismo, marketing y minería de texto.

Los artículos seleccionados para la revisión de la bibliografía se encuentran dentro del límite temporal de 5 años pertenecientes al intervalo de los años: 2015 – 2020.

CAPÍTULO IV: DESARROLLO Y RESULTADOS

4.1. ESTRUCTURA DE LA INFORMACIÓN

Durante el estudio de los artículos y trabajos de investigación utilizados en este proyecto, se tuvo que utilizar conocimientos previos relacionados a las diversas disciplinas de la ciencia de datos, ya que el análisis de sentimientos utilizando machine learning abarcan diversos temas que en los siguientes párrafos serán explicados haciendo uso de gráficos que demuestran la estructura y relación de estas diversas disciplinas.

Existen diversos campos que se encuentran dentro de la ciencia de datos, los cuales abarcan la gran variedad de datos que existen y utilizan, además, otras técnicas y herramientas que pertenecen a ciencias relacionadas. Algunos de estos son: Minería de Datos, Inteligencia Artificial, Big Data y las Matemáticas (Ver Figura 1).

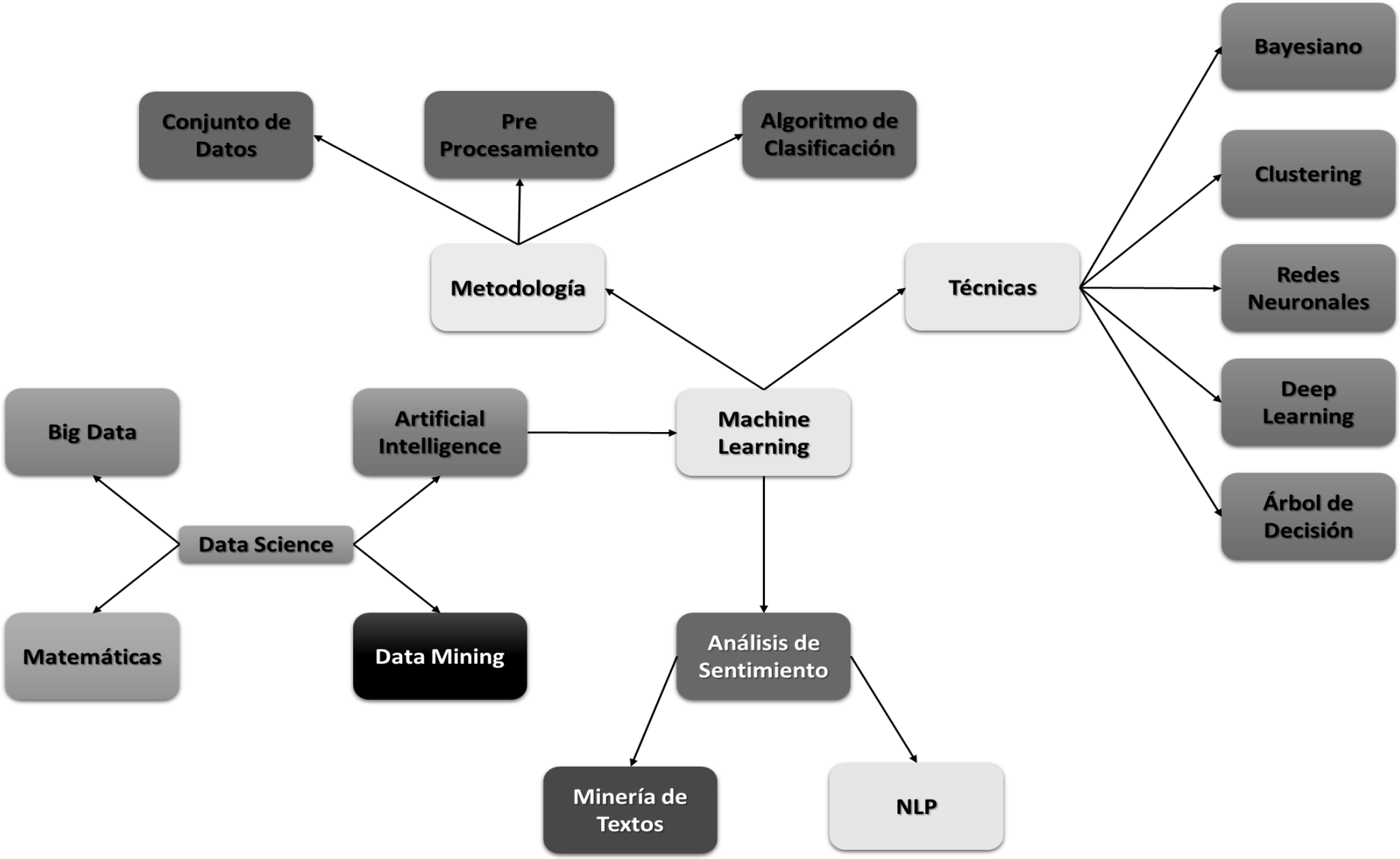

Figura 1: Estructura de la Información

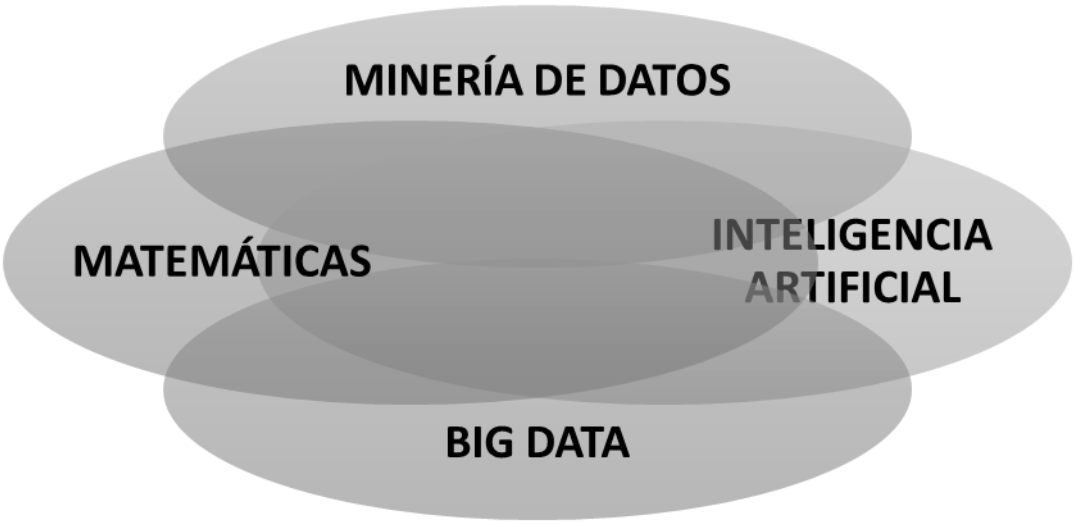

Figura 2: Campos en Data Science

De los campos anteriormente mencionados, se explicará a detalle el de la Inteligencia Artificial, en específico Aprendizaje automático, que contiene diversas técnicas para utilizadas para la clasificación de datos. Además, se mencionará la técnica de Procesamiento del Lenguaje Natural (NLP), debido a que es un campo que se enfoca a la comprensión del lenguaje humano mediante el ordenador. Esta, incorpora parte de la ciencia de datos, como el aprendizaje automático y la lingüística (Ver Figura 2).

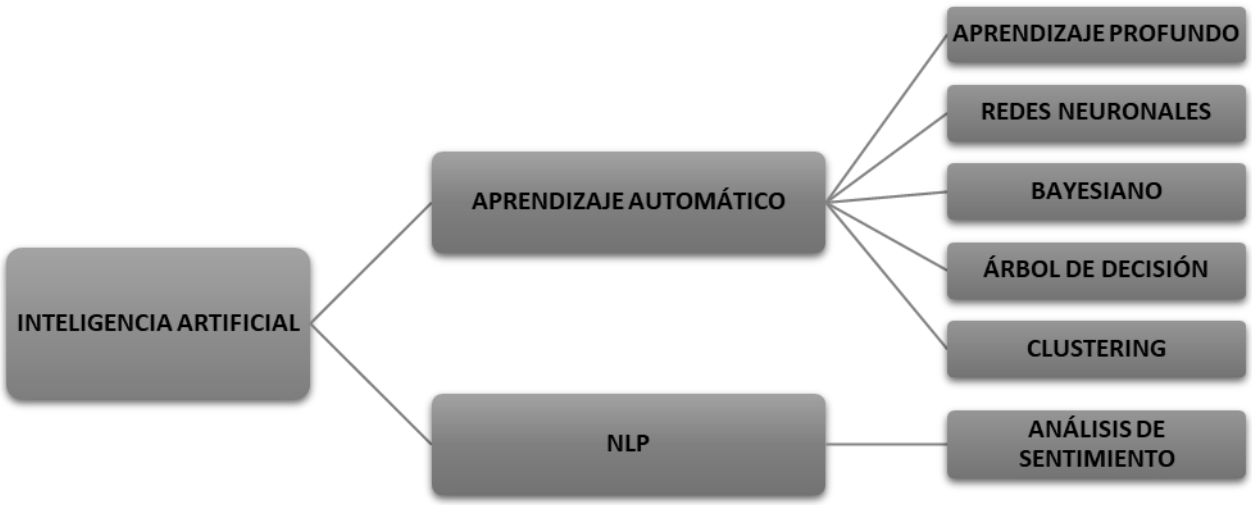

Figura 3: Técnicas de Machine Learning

4.2.PRINCIPALES HALLAZGOS DE ESTUDIO

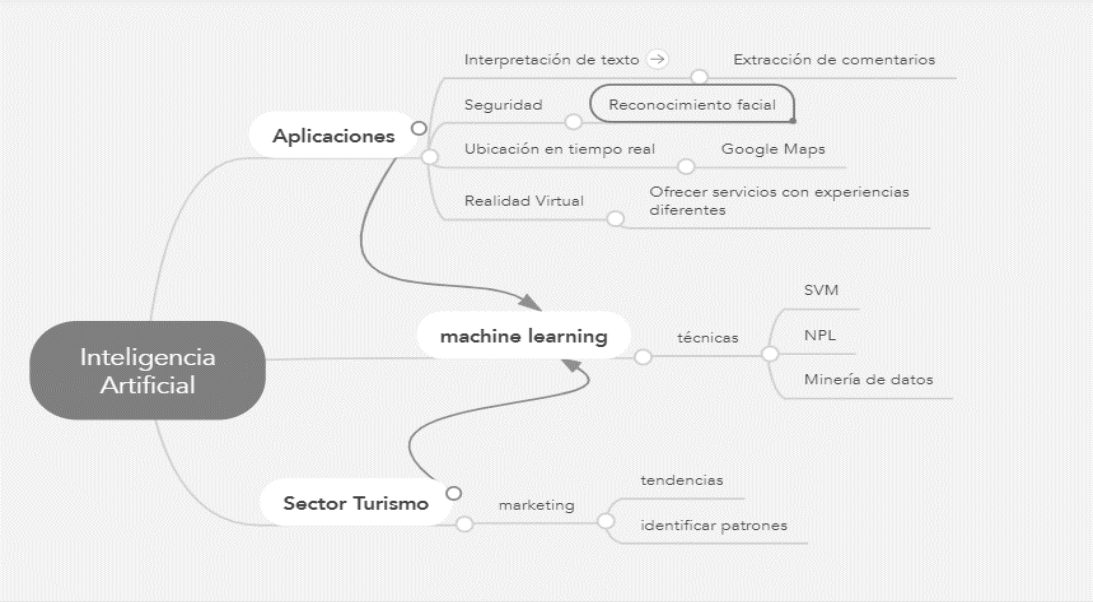

Figura 4: Mapa Mental principales hallazgos

Durante el proceso de búsqueda de información se ha observado que el campo de la inteligencia artificial ha ido cobrando mayor importancia en la vida cotidiana de las personas y las organizaciones.

El sector turístico, no es ajeno a esta tendencia, tal y como mencionan Samala, N., Shashanka, B., Shekhar, R., & Villamarin, R. (2020) en su artículo de investigación “Impact of AI and robotics in the tourism sector: a critical insight” cuyo objetivo es resaltar el rol de la inteligencia artificial en la industria del turismo. Menciona que esta tecnología es muy utilizada en todos los campos, logrando alcanzar mejores resultados de calidad de servicio. Esto se debe a la capacidad de que esta tecnología puede ser implementada en dispositivos móviles, siendo una realidad el reconocimiento facial, que puede aportar mucho en el tema de seguridad, tanto en los aeropuertos como hoteles o para encontrar a una persona a través de una imagen. Detección de textos para su traducción en otros idiomas, compartir ubicación en tiempo real, realidad aumentada para vivir una experiencia de una manera virtual. Todo esto, beneficia a las organizaciones que están en el sector turismo, porque genera mejor experiencia en los servicios que ofrecen a sus clientes,

lo cual puede tener un gran impacto económico de manera positiva.

Por otro lado, hoy en día unos de los medios de comunicación más usados para que las organizaciones interactúen con sus clientes se encuentran en las redes sociales, donde las empresas pueden realizar ofertas sobre sus servicios y saber de una manera que es lo que opinan sus clientes más concurrentes y así obtener una mejor información futura de los servicios que se puedan brindar para mejorar dichos productos y generar mayor atracción de clientes. Es por esto que uno de los hallazgos más importantes encontrados es con referencia al tema de análisis de sentimiento aplicado a los textos o comentarios que se puede encontrar en diferentes redes sociales. Liu, P., Nie, D., He, X., Zhang, W., Huang, Z., & He, K. (2019). En su artículo de investigación “Sentiment Analysis of Chinese Tourism Review based on Boosting and LSTM” cuyo objetivo es proponer un modelo LSTM basado en los comentarios de las personas. En esta investigación los autores mencionan que, una de las razones más importantes de los clientes al momento de querer realizar la adquisición de un servicio es el impacto de la revisión en línea que puedan realizar, es decir la búsqueda de las opiniones de otras personas sobre el servicio que estos quieren contratar. En el tema de turismo es algo puntual, ya que cada persona puede pensar u opinar diferente sobre algo entonces si la empresa turística no actúa de manera inmediata a estos comentarios puede verse comprometida de una forma negativa. Es por esta razón que los autores proponen un modelo basado en LSTM, según lo que indican es que, a diferencia de las demás técnicas de inteligencia artificial, esta puede aprender la dependencia de información a largo plazo y es capaz de minimizar el problema de clasificación que existen con respecto a los otros modelos.

Así mismo, se ha encontrado otras tecnologías que pueden realizar la misma tarea de clasificación de texto, Kirilenko, A., Stepchenkova, S., Kim, H., & Li, X. (2017). Realizaron la investigación “Automated Sentiment “. El cual tiene como objetivo evaluar la idoneidad de los diferentes tipos de clasificadores automáticos y comparar

el desempeño con el trabajo realizado por el humano. Donde mencionan un software llamado SentiStrength que realiza el análisis de sentimiento con mensajes cortos y/o frases, lo cual hace que hace que sea perfecto para calificar comentarios realizados en redes sociales.

Chen H. y Tang J. (2018) en su artículo científico “Web Evaluation Analysis of Tourism Destinations Based on Data Mining “, tienen como objetivo el análisis de las tendencias emocionales revisando diferentes publicaciones de turistas a través de OTA (Agencia de viajes Online). En su investigación se enfocaron en la industria del turismo en la ciudad Hainan, China. También indican que, es necesario realizar el análisis las necesidades de marketing de los lugares turísticos mediante la combinación del comportamiento de los consumidores en la web, analizando datos de consumo y tráfico, la integración de estos datos es posible utilizando Big Data. Para esta investigación, los autores realizaron un análisis emocional basado en la minería de texto web aplicando el modelo EVSM, que es utilizado para el cálculo de la relevancia de información y un algoritmo de clasificación de texto, logrando evaluar y clasificar las tendencias de los turistas obtenidos de la evaluación de OTA. Para este modelo se definen ciertos procesos claves, los cuales son importantes para lograr el objetivo final, que es el análisis de la información. Primero se recopila información utilizando herramientas de rastreo web, utilizaron como fuente de información la web del portal TripAdvisor. Segundo, construcción del diccionario emocional, esto con el fin de conocer las necesidades o maneras de pensar de los turistas de manera más efectiva. El diccionario se divide de acuerdo al léxico, biblioteca de adjetivos o biblioteca de verbos, para diferenciar el léxico utilizado en cada oración o frase. Tercero, preprocesado de texto, el objetivo que se busca en esta fase es extraer características en base a la frecuencia, a su vez se utiliza el algoritmo de indexación sobre el texto base de la matriz. Cuarto, Proceso de conocimiento según el patrón de texto, en esta etapa se adopta el algoritmo basado en el concepto, que es preparar el documento, construir el diccionario emocional de

los turistas, hacer un análisis estadístico, generar vectores de características. Finalmente, el método de evaluación del patrón del texto generará índices puntuales que deben seleccionarse bajo los criterios subjetivos y objetivos y en el enfoque cuantitativo para comparar el patrón efectivo, novedoso, disponible y comprensible en los conjuntos de patrones de resultados. En cuanto al análisis de los resultados, concluyeron que, el objeto en estudio, el caso de la ciudad de Hainan, China, fue evaluado positivamente con respecto a las referencias de los turistas sobre los servicios y/o diferentes características encontradas en su visita a dicho lugar. Por otro lado, se hizo un análisis de las palabras utilizadas en los cuestionarios, haciendo referencias como “hermoso paisaje” o “boletos con precios caros”. Los autores finalizan su investigación, mencionando que considerando que los textos encontrados en diferentes webs no son información estructurada, el método EVSM se utiliza para representar el conocimiento y apoyándose en algoritmos de clasificación se obtuvieron resultados positivos en esta investigación.

4.3. DESARROLLO DE LA INVESTIGACIÓN

4.3.1. MINERÍA DE TEXTOS

Es una de las técnicas utilizadas en la minería datos, por lo que adopta las técnicas de aprendizaje automático con el objetivo de reconocer patrones y comprender información nueva en grandes cantidades de texto no estructurado.

Algunos de los autores abarcan este tema dentro de sus investigaciones, como Eíto, R. y Senso, J. (2004), que realizaron el artículo de investigación "Minería textual", en el cual definen Text Mining como una tecnología que facilita la identificación y extracción conocimientos en base a documentos o cuerpos de texto, es decir, son actividades que extraen significado (datos estructurados) de un gran grupo de textos (datos no estructurada). Además, establecen desde una perspectiva comercial la relación entre Text Mining y Data Mining, ya que ambas utilizan grandes cantidades de información para extraer conocimiento, pero remarcan la diferencia, indicando que, Data Mining, a

comparación con Text Mining, utiliza patrones observables en datos estructurados y almacenados en bases de datos relacionales.

Por otro lado, el campo de la minería de texto se encuentra actualmente en auge debido a su importancia en el desarrollo de la tecnología, tal y como lo mencionan Srivatava, A. y Sahammi, M. (2013), que realizaron el libro de ciencias tecnológicas "Text Mining, Classification, Clustering, and Aplications", en el cual señalan el crecimiento de la información en los últimos años y su relación con la tecnología, ocasionando la complejidad de su análisis debido a la cantidad de información textual que se encuentra disponible en la web. En este contexto, se deriva la necesidad del uso de herramientas, y por ende la aparición de la minería de textos, como tecnología de uso frecuente para el acceso, análisis y procesamiento de grandes cantidades de información textual. Debido a sus beneficios que van mucho más allá de la búsqueda, la minería de texto se puede encontrar en aplicaciones basadas en el consumidor, así como en sistemas enfocados en la banca y finanzas, el cuidado de la salud, la industria aeroespacial, la fabricación y las ciencias naturales, ya que esta tecnología mejora la comprensión y el uso de la información en repositorios de documentos.

En conclusión, a lo mencionado por los autores, es una tecnología que se usa en distintas áreas, debido a su beneficio para la búsqueda y comprensión de textos utilizando información no estructurada, lo cual es fundamental hoy en día, por la cantidad y variedad de información que se almacena en uno de los instrumentos más importantes para las personas, la web.

4.3.2. ANÁLISIS DE SENTIMIENTO

Poblete, J. (2016), realizó el proyecto de investigación "Análisis de sentimiento y clasificación de texto mediante Adaboost Concurrente", en la cual define el análisis del sentimiento o minería de opiniones, como tareas relacionadas con el tratamiento computacional para la determinación de la actitud (juicio o evaluación) o la intención comunicativa emocional (intensión del autor para ocasionar emociones en el lector) de

un interlocutor o escritor, sobre un tema o contexto en un documento. Algunas de las tareas son: clasificación de documentos de opinión, extracción de representaciones estructuradas de opiniones, resumen de textos de opinión. entre otras.

Rincón, S. (2016), realizó el trabajo de investigación “Minería de textos y análisis de sentimientos en sanidadysalud.com”, en la cual detallan las tareas relacionadas a la minería de textos, siendo una de estas el análisis de sentimientos, con el objetivo de determinar la polaridad general en documentos para detectar la actitud del creador utilizando etiquetas como: positiva, negativa o neutra. Una de las diversas tareas más estudiadas dentro del análisis del sentimiento es la Clasificación de Sentimientos, utilizando un problema de clasificación con dos clases: positiva y negativa, en base a la subjetividad de las opiniones, para extraer información objetiva, por lo cual la clasificación es dependiente al dominio de las opiniones o documentos. También, se hace mención de la Web 2.0, como consecuencia del crecimiento de los medios de comunicación social en internet, siendo así fundamental desde la perspectiva empresarial, el uso de las herramientas de Análisis de Sentimientos para todo tipo de expresiones publicadas en la web (redes sociales, blogs, foros, etc.) como: opiniones de los productos, gestión de reputación e identificación de oportunidades de negocio, entre otros.

Rezza, A.; Becken, S. y Stantic. B. (2017), realizaron el artículo de investigación “Sentiment Analysis in Tourism: Capitalizing on Big Data”, en la que detallan los métodos de análisis de sentimientos que utilizarán para la ejecución de su proyecto. Definen el análisis de sentimiento como el uso de la lingüística computacional y el Procesamiento del Lenguaje Natural (PNL) para el análisis e identificación de información subjetiva en textos. Si bien las investigaciones relacionadas al análisis de sentimientos se remontan desde 1970, recién en las últimas décadas es que ha recibido atención por los investigadores y profesionales, debido a la escalada de información basada en web y redes sociales, la evolución de las tecnologías y especialmente enfoques de aprendizaje automático para análisis de textos, desarrollo de modelos de negocio y aplicaciones que

requieren del uso de este tipo de información. Debido a que las opiniones de clientes expresadas a través de reseñas de textos de publicaciones en las redes sociales a menudo son subjetivas u objetivas, metodológicamente representan un problema de clasificación de polaridad, por lo que se puede clasificar como binaria, ternaria u ordinal. Además, según investigaciones se considera importante la extracción de aspectos implícitos y explícitos en las revisiones para la precisión de los resultados de análisis de sentimientos. (p. 2-4)

4.3.3. PROCESAMIENTO DE LENGUAJE NATURAL (NLP)

Ticona R., (2019). en su tesis de investigación “Minería de opiniones basado en aprendizaje supervisado en la evaluación de destinos turísticos de la región de Puno”, definen NLP, citando diferentes autores, como Sosa E., (1997), quien señala que el NLP es la transformación del lenguaje humano utilizando sistemas informáticos. Por otro lado, mencionan a Liddy, E. (2001), quien sustenta que, el PLN son técnicas para el análisis y representación de texto naturales en diferentes niveles, para alcanzar el nivel similar del lenguaje humano para tareas o aplicaciones. (p.24)

En conclusión, el objetivo de NLP es procesar el lenguaje, brindando por el ser humano en texto, analizando las expresiones que se pueden encontrar utilizando aprendizaje automático.

4.3.4. MÉTODOS DE ANÁLISIS DE SENTIMIENTOS

4.3.4.1. APRENDIZAJE AUTOMÁTICO

Candia, D. (2019), realiza el trabajo de investigación “Predicción del rendimiento académico de los estudiantes de la UNSAAC a partir de sus datos de ingreso utilizando algoritmos de aprendizaje automático”, en el cual define el Aprendizaje Automático o Machine Learning como un conjunto de diversas diciplinas como Inteligencia Artificial, probabilidad y estadística, teoría de la información, entre otros. Mediante técnicas de Machine Learning, se permite a las máquinas aprender y comprender con información de situaciones actuales e históricos, utilizando algoritmos. El objetivo es lograr la

solución de problemas mediante la toma de decisiones utilizando experiencia acumulada. Mitchell en 1997 menciona que el aprendizaje automático abarca muchos campos, como la estadística, neurobiología, psicología, complejidad computacional, filosofía, teoría de la información y la teoría del control. En síntesis, el proceso de aprendizaje automático, depende del aprendizaje propio de la maquina mediante la selección de unas características importantes en un objeto de estudio y características conocidas realizando comparación entre ambas, para finalmente realizar la adaptación de estas, haciendo la vida de muchos, menos complicada (p. 16-18).

Así mismo, el aprendizaje automático está basada en utilizar experiencias pasadas que sirvan a la máquina para poder alcanzar un nivel óptimo de aprendizaje llegando a predecir eventos futuros en base a los patrones de los modelos obtenidos. Alpaydin, E. (2010), realizó el trabajo de investigación “Introduction to Machine Learning”, en el cual señala que el aprendizaje automático es una programación de computadoras para optimizar un criterio de rendimiento utilizando datos de ejemplo o experiencias pasadas. Esto quiere decir que este aprendizaje depende mucho de la información que ha recibido previamente y también depende de las veces que se ha entrenado para alcanzar un nivel óptimo muy alto y llegando a obtener buenos resultados. También, está basado en la teoría de la estadística porque aplica la inferencia a partir de una muestra, con el fin de evaluar las diferentes posibilidades de alcanzar un resultado.

4.3.4.1.1. TIPOS DE APRENDIZAJE

a) APRENDIZAJE SUPERVISADO

Rodriguez, J. y Minaño, M. (2017) elabora el proyecto de tesis “Desarrollo de una aplicación informática basada en un modelo de Machine Learning para mejorar la evaluación de préstamos crediticios”, en el cual citan la definición de Aprendizaje Supervisado, dada por Alpaydin (2010), como una modalidad que se da a través de un algoritmo, donde se le determinan ciertas preguntas y respuestas, para lo cual se realizan etiquetas para posibles respuestas y

etiquetas para posibles preguntas. Todo ello con el fin de comprender la entrada de toda la información y salida, para que cuando al sistema se le formulen preguntas similares (utilizando etiquetas) sepa que responder mediante predicciones.

Ticona, R. (2019), realiza el informe de investigación “Minería de opiniones basados en aprendizaje supervisado en la evaluación de destinos turísticos de la región de Puno”, en el cual menciona que, en los sistemas de aprendizaje supervisado, es importante clasificar correctamente la información como por ejemplo positiva o negativa, con el fin de sumarle al documento o información un emblema que ayude a asociar datos desconocidos, pero que, sin embargo, el sistema gracias a la clasificación pueda asociarlo a alguna similar aumentando la probabilidad de acierto.

En conclusión, ambos autores hacen mención que al utilizar la técnica de Aprendizaje Supervisado se requiere de un entrenamiento previo para que el algoritmo aprenda con datos etiquetados, es decir, respuestas conocidas, los cuales servirán como guía para posteriores respuestas a resultados de datos desconocidos.

b) APRENDIZAJE NO SUPERVISADO

Florián, J. (2013) en su tesis de pregrado “Categorización de texto usando técnicas de Machine Learning aplicado a la clasificación de reclamos en los procesos de la Universidad Tecnológica de Bolívar”, señala que en la técnica de Aprendizaje No Supervisado, no existe un proceso guía del aprendizaje, por lo cual se usa una medida de los datos para que el algoritmo aprenda, y se optimice parámetros con respecto a esta medida. En base a esto, se forman representaciones internas de características correspondientes a los datos que se usan de entrada, para que finalmente se creen nuevas clases de forma automática.

Candia, D. (2019), realiza el trabajo de investigación “Predicción del rendimiento académico de los estudiantes de la UNSAAC a partir de sus datos de ingreso utilizando algoritmos de aprendizaje automático”, en el cual cita a varios autores para definir el Aprendizaje No Supervisado, uno de ellos es, Palaez, N. (2012) señaló que el aprendizaje no supervisado solo abarcaba el tema de correlaciones o características dentro de la red, categorizando los datos, con el propósito de asociar parecido entre sí mismos. Calvo, D. (2017) indica que este aprendizaje son una mezcla de técnicas que hacen que puedan inferir modelos para obtener respuesta de datos desconocidos. Por otro lado, Gonzalo, A. (2018) define que este sistema es subjetivo ya que no brinda ni tiene la capacidad de dar respuestas correctas, solo sobreentender. El considera que el objetivo del ANS es distribuir correctamente la información para aprender de ello, resumiendo lo esencial.

4.3.4.1.2. TÉCNICAS DE APRENDIZAJE AUTOMÁTICO

Entre las diferentes técnicas existentes de machine learning utilizados en modelos de clasificación de estudios de investigación, que han permitido automatizar la clasificación de opiniones según el sentimiento que expresan, destacan los siguientes los modelos mayormente empleados: Máquina de Vectores de Soporte (SVM), Nave Bayes, K-NN, Random Forest, Árbol de decisiones y Aprendizaje Automático Extremo (ELM) (Ver Tabla 2).

Tabla 2: Términos de clasificadores

Clasificador	Descripción
K-NN	Clasificación no paramétrica, basado en instancias. Calcula la distancia de una muestra con la de todo el resto, examina su relación con los otros grupos mediante k observaciones, para asignarlo al grupo con relación

	a esas observaciones.
SVM	Algoritmo de aprendizaje automático supervisado, que utiliza un hiperplano de separación para la clasificación de datos. Dicho hiperplano, requiere de entrenamiento previo usando etiquetas de forma que segregue los datos de manera óptima.
Naïve Bayes	Funciona en base a suposiciones de que las características de los datos son todas independientes.
Árbol de decisión	Es un diagrama en forma de árbol, cuyas ramas representan una serie de condiciones que concurren secuencialmente para la resolución de un problema.
ELM	Es una red neuronal que no requiere de proceso de aprendizaje previo para obtener parámetros de los modelos.
Random Forest	Es un algoritmo de clasificación que refiere a un conjunto de árboles de decisiones no correlacionales, agrupados según la aleatoriedad de características de cada árbol individual.

Para comprender los algoritmos de machine learning mencionados anteriormente, en los siguientes párrafos se realizará una descripción de algunos, lo cual permitirá un mayor entendimiento durante la comparativa de modelos empleados en los artículos de investigación seleccionados.

a) SVM

Sharma, S. y Sharma, V. en su investigación “Performance of Various Machine Learning Classifiers on Small Datasets with Varying Dimensionalities: A Study”, definen la Máquina de Vectores de Soporte como clasificadores basados en el principio de la minimización de riesgo estructural y en la teoría del aprendizaje estadístico, cuyo objetivo es la determinación de hiperplanos (límites de decisión) para poder separar las clases de forma

eficiente. Aunque los SVM presentan lentitud en su entrenamiento, la presión con la que modela límites de decisión complejos, conlleva a que sean poco propensos ante ajustes excesivos a diferencia de otros métodos.

Así mismo, según la eficiencia de este modelo de aprendizaje, Rezza, A.; Becken, S. y Stantic. B. (2017), en su artículo de investigación “Sentiment Analysis in Tourism: Capitalizing on Big Data”, indican que el clasificador SVM y Naïve Bayes son los métodos claves del aprendizaje automático utilizados para los análisis de sentimientos en la literatura, ya que su diseño es específico para problemas que presentes dos tipos de clases. También señalan que el SVM utiliza datos anotados previamente, por lo cual se encuentra dentro del enfoque de aprendizaje supervisado, ya que para obtener un hiperplano que clasifique con precisión los datos de nuevas muestras en grupos, se requiere un previo entrenamiento con datos conocidos, es decir etiquetados (Ver Figura 3). Sin embargo, este método requiere de menos datos con anotaciones en comparación con enfoques basados en redes neuronales.

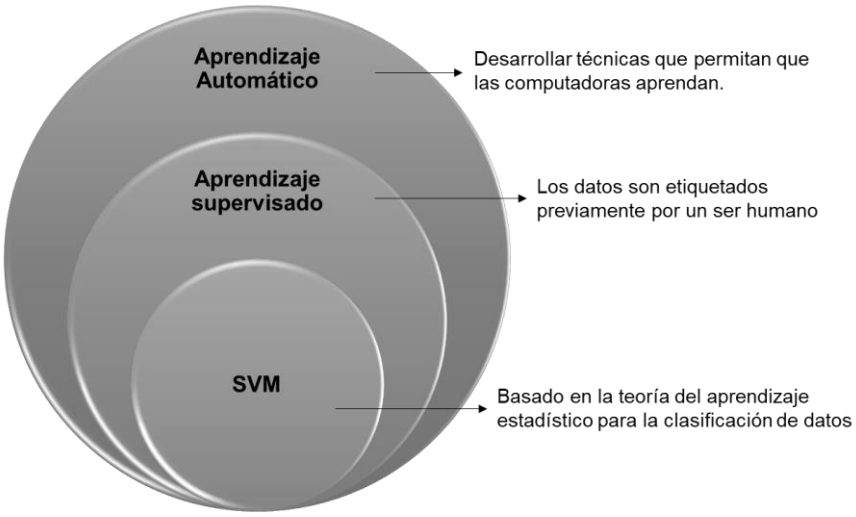

Figura 5: SVM - Aprendizaje Supervisado:

En conclusión, para el desarrollo del proyecto uno de los clasificadores más aptos y estudiados es el SVM, debido al tratamiento de datos

textuales dentro del Análisis de Sentimiento. Su uso requiere de hiperplanos (Ver Figura 4), las cuales resultan de una ecuación, para clasificar datos, sin embargo, pueden existir diversos hiperplanos posibles, por lo el objetivo de este algoritmo es encontrar el que esté a una mayor distancia de los datos de las clases.

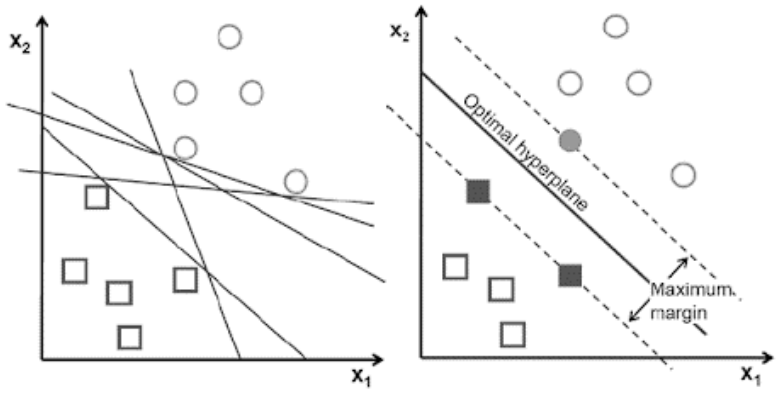

Figura 6: Hiperplano para 2D

Además, si el espacio que existe entre los dos conjuntos de datos no se puede separar linealmente, es decir, no existe un hiperplano que los separe para distinguir las clases, es posible utilizar una de las funciones del núcleo (Kernel). La cual tiene diferentes tipos: Polinomial-Homogénea (Ver Figura 5), Base Radial Gaussiana (Ver Figura 6) y Perceptron (Ver Figura 7).

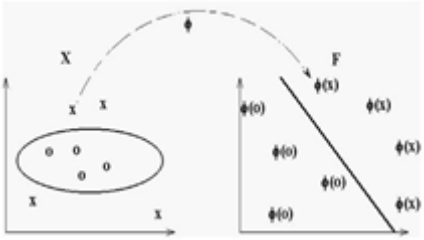

Figura 7: Kernel Polinomial

Ecuación: $K(xi, xj) = (xi \cdot xj)n$

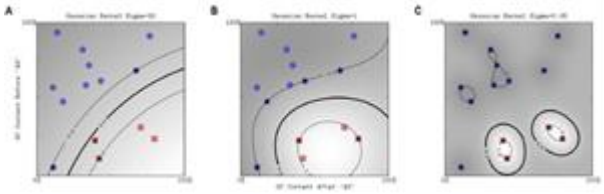

Figura 8: Base Radial Gaussiana

Ecuación: $K(xi, xj)=exp(-(xi-xj)2/2(sigma)2)$

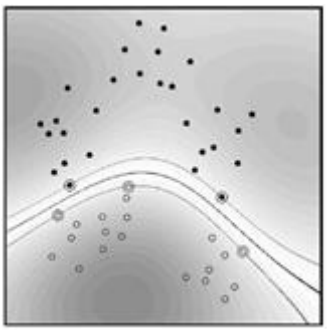

Figura 9: Perceptron

Ecuación: $K(xi, xj)= || xi-xj ||$

b) NAÏVE BAYES

Cárdenas, J.; Olivares, G. y Alfaro, R. en su artículo de investigación "Clasificación automática de textos usando redes de palabras", definen Naïve Bayes, citando a varios autores para analizar el desempeño de este clasificador de textos, entre los cuales está Lewis, D. (1998), quien señala que este algoritmo es un modelo estructural y utiliza un conjunto de probabilidades para determinar la relación entre los nodos(atributos) y los enlaces(dependencias). Por otro lado, Zhang, H. (2004), aporta fundamentos positivos de clasificación de problemas reales utilizando este método. Para lo cual, Caruana, R. y Niculescu-mizil, A. (2006), realizan un trabajo de investigación en el cual demuestran que el desempeño de este algoritmo es inferior con respecto a la utilización de otros métodos.

Por otro lado, Ticona, R. (2019), en su informe de investigación "Minería de opiniones basados en aprendizaje supervisado en la evaluación de destinos

turísticos de la región de Puno", utiliza el clasificador Naïve Bayes para el análisis de sentimiento de comentarios en Twitter, por lo cual en su marco teórico realiza la definición de este método como un clasificador supervisado y generativo, basado en el Teorema de Bayes y en la independencia de cada atributo perteneciente a una clase. Se le llama Naïve (ingenua) debido a que el concepto del método difiere con la realidad, es decir, los atributos en la práctica mayormente presentan dependencia entre sí.

4.3.5. METODOLOGÍA DE ANÁLISIS DE SENTIMIENTOS CON MACHINE LEARNING

Para el desarrollo de la investigación se planteará un esquema relacionado a las metodologías realizados en las distintas investigaciones seleccionadas (Ver Figura 8). A partir de ello, se realizará una comparación de los diferentes conjuntos de datos utilizados como fuente de información, técnicas de preprocesamiento de texto y algoritmos para la clasificación de texto, comentarios u opiniones, que servirá para la identificación de los modelos más óptimos empleados, basados en su rendimiento y precisión.

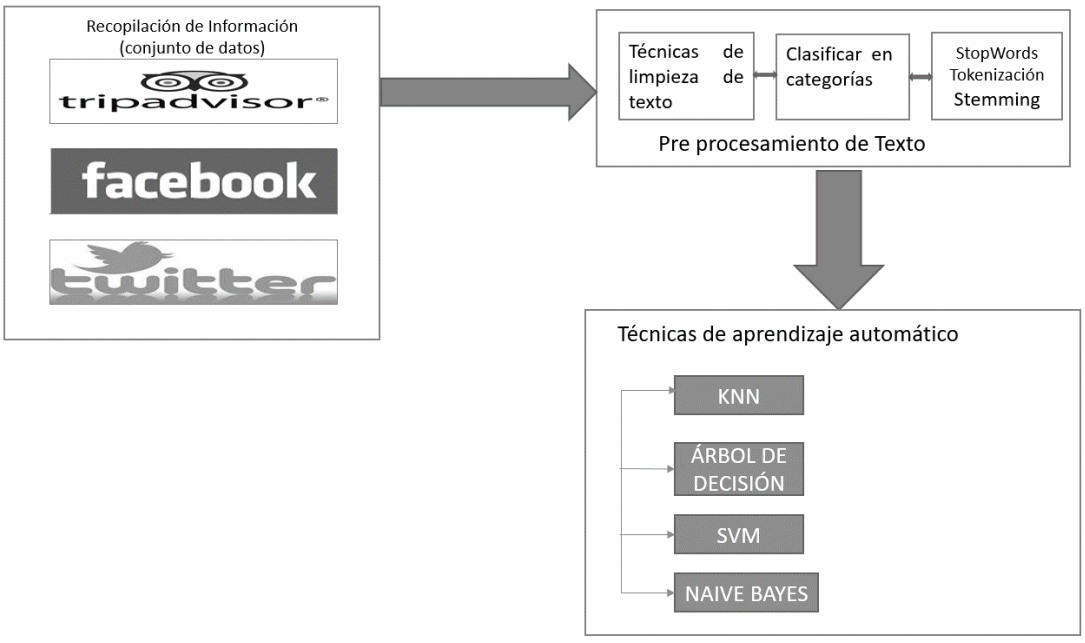

Figura 10: Esquema de las metodologías según las distintas fuentes bibliográficas

4.3.5.1. CONJUNTO DE DATOS

La fuente de información de datos sirve como valores de entrada para los algoritmos de clasificación, las cuales deben ser precisas y reales para poder ser empleada de una manera correcta. Por esta razón, Thelwall, M. (2019) en su trabajo de investigación “Sentiment Analysis for Tourism”. Menciona que la recopilación de comentarios se obtuvo a través del portal web TripAdvisor y también hace mención a las redes sociales que son una fuente de información amplia, ya que la mayoría de personas pueden escribir lo que piensan por medio de estos sitios de internet. Asimismo, Medeiros, M., Silveira, D., Fernandes, L. y Soares, A. (2018). Realizaron el artículo de investigación “Un análisis a partir del contenido generado por los usuarios en Tripadvisor”. Ellos indican que, el sitio web más resaltante para la extracción de comentarios u opiniones es TripAdvisor, porque es un sitio web confiable y con diversa información de usuarios de todas partes del mundo con respecto a temas hoteles, lugares turísticos, restaurantes, etc. Por otra parte, Mridula A. Kavitha C. (2018). Realizaron la investigación “Estudio de opinión sobre minería y opinión de tweets Polaridad con el aprendizaje automático “. Describen que el conjunto de datos ideal se puede extraer de la red Social Twitter, porque cuenta con un API que sirve para la extracción de comentarios en tiempo real y además que los comentarios pueden tener un máximo de 280 caracteres. Sin embargo, Gao, S., Hao, J., y Fu, Y. (2015) desarrollaron el trabajo de investigación "The application and comparison of web services for sentiment analysis in tourism", recopilaron datos la página de TripAdvisor. No obstante, Prerna, M., Ranjana, R. y Pankaj, K. (2016). desarrollaron el trabajo de investigación "Sentiment Analysis of Twitter Data:Case Study on Digital India", para lo cual basaron la recolección de la información en comentarios extraídos de la red Twitter.

Como indican los autores mencionados anteriormente, resaltan que TripAdvisor, Twitter y otras redes sociales suelen ser las fuentes de información que son más utilizadas para

extraer comentarios, opiniones o textos y sobre esto aplicar un análisis de sentimiento a través de los diferentes método o algoritmos que existen. Esto se debe, a la cantidad de usuarios que tienen acceso a estos sitios web o redes sociales. Se puede apreciar que la mayoría de investigaciones realizadas utilizan las redes sociales, porque, es la manera digital de estar más cerca y en contacto con los diversos usuarios y sus opiniones. También, al ser un servicio publicado en una red de internet, todos pueden tener acceso, lo cual hace fácil su recopilación y poder analizarla utilizando técnicas de aprendizaje o automático, logrando así mejorar el proceso de marketing, tomando mejores decisiones y creando tendencias o estrategias personalizadas hacia los usuarios para alcanzar mayor número de clientes.

4.3.5.2. PRE PROCESAMIENTO DE TEXTO

En este punto, analizaremos uno de los pasos más importantes para la clasificación de texto, el cual consistente en procesar previamente la información obtenida desde los diferentes conjuntos de datos. Por ello, Mridula A. Kavitha C. (2018). Realizaron la investigación “Estudio de opinión sobre minería y opinión de tweets Polaridad con el aprendizaje automático “. Para el preprocesamiento de texto, emplearon técnicas de limpieza de texto, para evitar palabras que no existen y poder agrupar los comentarios de acuerdo a su polaridad, es decir, negativo, positivo y neutro, llamada también como fase labeling o etiquetado en español. Con ello, lo que buscan es que los datos de entrada para el algoritmo, sean lo más preciso posible para lograr obtener mejores resultados. Así mismo, Manchado, F. (2018) en su trabajo de investigación "Análisis de sentimientos basado en opiniones turísticas", plantea que, para el Procesamiento de Lenguaje Natural, es necesario aplicar Word2Vec, ya que, son modelos aplicados para vinculación de palabras o frases a vectores numéricos, el objetivo de esto es la obtención de los vectores de características para luego poder ser clasificados. Por otro lado, Amaya, C., Magaña, P., y Ochoa, I. (2017) desarrollaron el trabajo de investigación "Evaluación de destinos turísticos mediante la tecnología de la ciencia de datos". Los

autores plantean, que después de la extracción de los datos, estos se almacenan en un repositorio para previamente clasificar las categorías en las cuales puedan ser seleccionadas, posteriormente aplicando LNP para clasificar los sentimientos. Por otro lado, Prerna, M., Ranjana, R. y Pankaj, K. (2016). desarrollaron el trabajo de investigación "Sentiment Analysis of Twitter Data:Case Study on Digital India", para el preprocesamiento utilizaron módulos de NLTK, conjunto de programas para el procesamiento de lenguaje natural simbólico en el entorno Python, el cual incluye las tareas de StopWords, Tokenización y Stemming. Todo esto con el fin de poder tener textos claros y limpios para realizar un análisis correcto y preciso.

Existen diversas técnicas de pre procesamiento de texto y posteriormente aplicar el análisis del mismo. Sin embargo, en base a la revisión bibliográfica, la técnica de procesamiento que podría dar una mayor precisión para realizar una correcta clasificación estaría referida con respecto a lo que mencionan los autores Prerna, Ranjana y Pankaj, utilizar módulos de NLTK, ya que es más completa y tiene más procesos para la limpieza y clasificación de texto. Tales como StopWords, que sirve para extraer palabras que suelen ser vacías o que no existen, luego el proceso de Tokenización, que consiste en separar las palabras de un texto y construir en un vector compuesto por cada palabra que forma parte de dicho texto, Finalmente el método de Stemming, es un método que reduce su palabra original en la parte raíz de la misma, consiguiendo disminuir el grupo de características con el fin de facilitar la clasificación posterior.

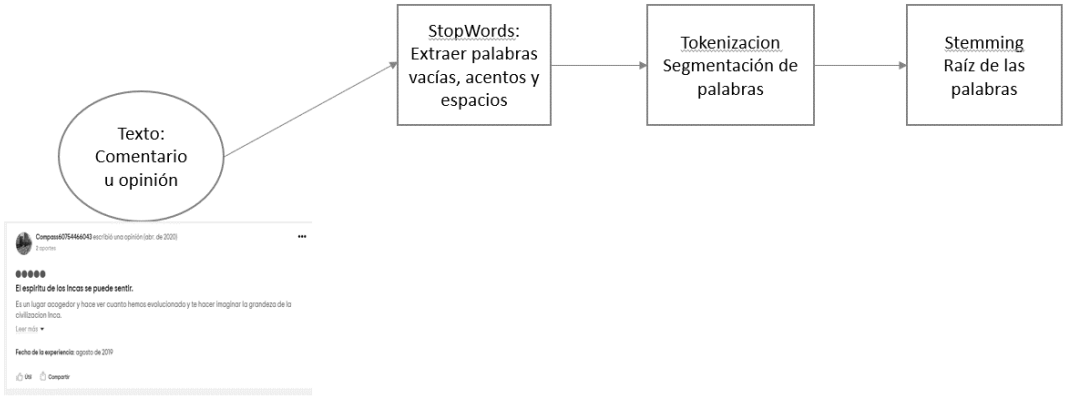

Figura 11: Preprocesamiento utilizando módulos NLTK

4.3.5.3. ALGORITMOS DE CLASIFICACIÓN DE TEXTO

Existen diversas investigaciones desarrolladas por los intereses de académicos, investigadores o empresarios, que buscan poder hacer uso de la gran cantidad de información que se encuentran disponibles en Internet, relacionada a opiniones, sentimientos y emociones que los consumidores expresan hacia determinados productos y servicios, con la finalidad de clasificar dichos conjuntos de datos según la polaridad de los sentimientos, sean positivos o negativos, que expresa cada texto. Con ello, se intenta solucionar el problema que las organizaciones enfrentan durante la toma de decisiones con respecto a la satisfacción de sus consumidores.

Hay diversos enfoques para el análisis de sentimientos, por lo cual Thelwall, M. (2019) realizo la investigación “Sentiment Analysis for Tourism”, con el objetivo de revisar acerca de los principales enfoques de análisis de sentimientos describiendo como se aplican y como es el funcionamiento de los métodos. Como primer enfoque menciona el análisis del sentimiento léxico, utilizando un diccionario de términos relacionados con sentimientos asignando una puntuación de acuerdo a su polaridad y promedio en uso común. Como segundo enfoque, hace referencia al análisis de sentimiento de aprendizaje automático, esto se puede lograr sin un léxico utilizando esta nueva tecnología (Liu, 2012), ya que generalmente se enfoca en las palabras y no en las

oraciones completas. Realizando una comparativa entre ambos enfoques, se concluyó que el aprendizaje automático para el análisis de sentimiento genera muchas más ventajas, ya que se puede analizar grandes cantidades de texto, los cuales no requieren de esfuerzo humano, para volver al sistema cada vez más preciso con la información recibida, que puede ser tomada por diferentes sitios web.

Así mismo, diversos autores utilizan técnicas de machine learning para la clasificación de opiniones. Uno de ellos es Mridula A. y Kavitha C. (2018), que realizaron la investigación “Estudio de opinión sobre minería y opinión de tweets Polaridad con el aprendizaje automático “, con el objetivo de analizar y encontrar el algoritmo más preciso para el caso en estudio empleando los clasifidores probabilísticos de aprendizaje automático supervisado Naive Bayes y SVM. El análisis de los resultados se aplicó a 29 mil tweets filtrados por el límite temporal de un mes y que incluyan el hashtag #MakeIndia. Se obtuvo que ambos algoritmos en cuanto a la precisión su rendimiento es bajo. En dicho estudio, ninguno de los clasificadores tiene precisiones altas. Sin embargo, SVM supera a Naïve Bayes en un 8%.

Al igual que el artículo de investigación mencionado anteriormente, Kirilenko, A., Stepchenkova, S., Kim, H., & Li, X. (2017), realizaron la investigación “Automated Sentiment “, en el cual utilizaron Naive Bayes y SVM, con el objetivo evaluar la capacidad de estos clasificadores automáticos y comparar el desempeño con el trabajo realizado por el humano. De los resultados se obtuvieron que, los algoritmos de aprendizaje SVM y Naive Bayes alcanzaron un nivel similar al de los evaluadores con una precisión de 0.89 y 0.85, respectivamente.

Con ello, se observa que la mayoría de investigaciones utilizan técnicas de aprendizaje automático, siendo las más populares SVM y Naive Bayes, que, en comparativa entre las investigaciones mencionadas, SVM supera en presión de clasificación de opiniones, al otro clasificador. Sin embargo, estos no son las únicas técnicas empleadas. Existen otras investigaciones que abarcan una comparación más extensa de técnicas, como la que se mencionará a continuación.

Los autores Topal, I. y Muhammed K. (2016) realizaron el artículo de investigación “Automatic Determination of Travel Preferences of Chinese Tourists” que tiene como objetivo analizar datos históricos de los usuarios chinos de TripAdvisor con métodos de inteligencia artificial para revelar el perfil del consumidor que podrían preferir visitar Turquía o Francia. Para lo cual, utilizaron 5 métodos de clasificación distintos los cuales son: Árbol de decisión, K vecinos más cercanos (KNN), redes neuronales artificiales de avance multicapa (MLFFNN), Redes neuronales probabilísticas (PNN) y máquinas de soporte de vectores (SVM). Además, se utilizó el Clasificador de conjunto, el cual engloba todos los 5 mencionados anteriormente. De los resultados se obtuvo que, la mejor sensibilidad para los usuarios chinos que eligieron Turquía se obtuvo como 1 con KNN y para Francia 0.94 con MLFFNN. Además, se observó que a medida que aumentaba el número de características en los conjuntos de datos el rendimiento aumentaba.

CAPÍTULO V: CONCLUSIONES

5.1. TENDENCIAS

En el presente capitulo, se detallarán las conclusiones en base a los artículos de investigación utilizados como referencia bibliográfica anteriormente. Con ello, se pretende aportar a futuras investigaciones que realicen estudios enfocados en el proceso de marketing digital en el sector turismo empleando técnicas de aprendizaje automático.

En las últimas décadas los avances tecnológicos han permitido que más personas se sumen a la nueva era digital. En la actualidad, se ha evidenciado que la mayor fuente

de información de opiniones de las personas son las redes sociales, en donde se puede compartir datos de todo tipo desde fotos hasta opiniones personales con respecto a un tema en específico. Logrando así obtener el interés de las empresas para optar por mejores decisiones en base a la información que el usuario expresa con respecto a los servicios o productos que las organizaciones ofrecen. A partir de ello, nace la necesidad de realizar investigaciones que permitan realizar el análisis de opiniones alojados en los diferentes sitios web. Siendo el sector turismo uno de los interesados en la aplicación de las técnicas de machine learning para mejorar sus procesos de marketing en relación al análisis de sentimientos. Con ello, se espera optimizar la toma de decisiones que le permitirán a las empresas conocer las necesidades de sus clientes y mejorar sus productos o servicios para su crecimiento empresarial.

Con referencia al desarrollo de la investigación, se ha encontrado diferentes técnicas de aprendizaje automático como herramienta de apoyo para mejorar y agilizar los procesos de las organizaciones. Para alcanzar una correcta aplicación de esta tecnología es importante que se utilice de manera apropiada tanto la extracción como la recopilación de información, además de una buena limpieza de la misma y en el momento que sea procesada por un clasificador alcance niveles óptimos y la información obtenida sea precisa y no errónea.

5.2. ENCUENTROS Y DESENCUENTROS ENTRE LOS ESTUDIOS

Durante el análisis realizado para las distintas técnicas de modelos de aprendizaje automático, se llega a la conclusión que, para las etapas de la metodología planteada, que son conjunto de datos, pre procesamiento de datos y algoritmos de clasificación, se observó que:

- Para el conjunto de datos, se puede extraer las opiniones de los usuarios de diferentes plataformas web, siendo la que más resalta TripAdvisor, debido a que está enfocada principalmente comentarios sobre el sector turismo, ya sea, restaurantes, hoteles o destinos turísticos. Sin embargo, otros autores prefieren a las redes sociales, como Facebook, Twitter, porque, son más

utilizadas por los usuarios de diferentes partes del mundo y son más accesibles.

- Para la etapa de preprocesamiento de datos, algunos autores describen que, para una correcta clasificación, se debe realizar una limpieza de texto, extrayendo palabras innecesarias, agrupando en una misma clase de acuerdo a su polaridad. Por otro parte, hay otros autores mencionan a Word2Vec, ya que, son modelos aplicados para la producción de palabras o frases a vectores numéricos, el objetivo de esto es la obtención de los vectores de características para luego poder ser clasificados.
- Los algoritmos SVM y Naive Bayes son comúnmente utilizados para la clasificación de sentimiento de opiniones públicas en la web, debido a la precisión de sus resultados que aproximan a 0.9 de aciertos. Sin embargo, comparados con KNN estás mejoran su rendimiento teniendo como entrada grandes cantidades de características.

5.3. RESPONDE A LA PREGUNTA DE INVESTIGACIÓN

Con respecto a la pregunta de investigación planteada: ¿Qué modelos de aprendizaje automático optimizan el proceso de marketing digital en el sector turismo? Se responde en base al análisis de todas las tecnologías empleadas en los distintos trabajos de investigación, para lo cual, en la etapa de definición de conjunto de datos, se puede decir que la red social Twitter, tiene la ventaja, ya que, cuenta con un API que puede funcionar en tiempo real, ayudando a extraer comentarios en línea para luego ser preprocesados por otros métodos. Sin embargo, el más utilizado es TripAdvisor porque está netamente enfocado en temas relacionados al sector turismo.

Una vez obtenido el conjunto de datos de la fuente seleccionada, se procede a preprocesar esta información con diferentes técnicas. La técnica más recomendada es NLKT, ya que cuenta con diferentes bibliotecas para procesar texto, siendo más completa que las demás y funciona en la plataforma de Python, siendo de código abierto

y gratis.

Para la clasificación de los sentimientos de los datos obtenidos de las opiniones preprocesadas, se utilizan diversos tipos de algoritmos de machine learning, debido a la precisión y rendimiento para el análisis de texto. Según las revisiones realizadas a los artículos de investigación, comúnmente se utiliza el algoritmo SVM y Naive Bayes, comparándolos y obteniendo en algunos casos resultados superiores en SVM. Sin embargo, cuando se utilizan grandes cantidades de datos como entrenamiento y prueba, KNN tiene mejor rendimiento.

Por ello, para el uso de las técnicas en investigaciones relacionadas a la clasificación de opiniones según el sentimiento que expresan, debe considerarse una serie de factores, como el uso de palabras coloquiales o jergas en las opiniones deben eliminarse ya que dificultan la clasificación y disminuyen el rendimiento de la red.

CAPÍTULO VI: REFERENCIAS

6.1. REFERENCIAS BIBLIOGRÁFICAS

Afzaal, M., Usman, M., Fong, A., Fong, S., & Zhuang, Y. (2016). Fuzzy Aspect Based Opinion Classification System for Mining Tourist Reviews. *Hindawi*, 1-14. doi:10.1155/2016/6965725

Alaei, A., Becken, S., & Stantic, B. (2017). Sentiment Analysis in Tourism: Capitalizing on Big Data. *Journal of Travel Research, 58*, 175-191. doi:10.1177/0047287517747753

Amaya, C., Magaña, P., & Ochoa, I. (2017). Evaluación de destinos turísticos mediante la tecnología de la ciencia de datos. *Estudios y Perspectivas en Turismo, 26*, 286-305. Obtenido de https://dialnet.unirioja.es/servlet/articulo?codigo=6327726

Bucur, C. (2015). Using Opinion Mining Techniques in Tourism. *Procedia Computer Science, 23*, 1666-1673. doi:10.1016/S2212-5671(15)00471-2

Chen, H., Tang, T., Dai, Z., & Li, M. (2018). Web Evaluation Analysis of Tourism Destinations Based on Data Mining. *International Conference on Computer and Communications*, 1803-1808. doi:10.1109/CompComm.2018.8781024

Da'u, A., Salim, N., Rabiu, I., & Osman, A. (2019). Recommendation system exploiting aspect-based opinion mining with deep learning method. *Information Sciences*. doi:10.1016/j.ins.2019.10.038

Gao, S., Hao, J., & Fu, Y. (2015). The application and comparison of web services for sentiment analysis in tourism. *IEEE*. doi:10.1109/ICSSSM.2015.7170341

Gao, X., Tan, R., & Li, G. (2020). Research on Text Mining of Material Science Based on Natural Language Processing. *IOP Conference Series: Materials Science and*

Engineering, 768. doi:10.1088/1757-899X/768/7/072094
Giglio, S., Bertacchini, F., Bilotta, E., & Pantano, P. (2019). Using social media to identify tourism attractiveness in six Italian cities. *Tourism Management, 72*, 306-312. doi:10.1016/j.tourman.2018.12.007
Giglio, S., Bilotta, E., Pantano, P., & Bertacchini, F. (2019). Machine learning and point of interests: typical tourist Italian cities. *Current Issues in Tourism*, 1-13. doi:10.1080/13683500.2019.1637827
Kirilenko, A., Stepchenkova, S., Kim, H., & Li, X. (2017). Automated Sentiment Analysis in Tourism: Comparison of Approaches. *Journal of Travel Research, 57*, 1012-1025. doi:10.1177/0047287517729757
Li, J., Xu, L., Tang, L., Wang, S., & Li, L. (2018). Big data in tourism research: A literature review. *Tourism Management, 68*, 301-323. doi:10.1016/J.TOURMAN.2018.03.009
Li, W., Guo, K., Shi, Y., Zhu, L., & Zheng, Y. (2017). Improved New Word Detection Method Used in Tourism Field. *Procedia Computer Science, 108*, 1251-1260. doi:10.1016/j.procs.2017.05.022
Liu, P., Nie, D., He, X., Zhang, W., Huang, Z., & He, K. (2019). Sentiment Analysis of Chinese Tourism Review based on Boosting and LSTM. *International Conference on Communications, Information System and Computer Engineering (CISCE)*, 664-668. doi:10.1109/CISCE.2019.00154
Machado, F. (2018). Análisis de sentimientos basado en opiniones turísticas. *Universidad de La Laguna*. Obtenido de https://riull.ull.es/xmlui/bitstream/handle/915/10412/Analisis%20de%20sentimientos%20basado%20en%20opiniones%20turisticas..pdf?sequence=1
Martín, C., Torres, J., Aguilar, R., & Díaz, S. (2018). Using Deep Learning to Predict Sentiments: Case Study in Tourism. 1-9. doi:10.1155/2018/7408431
Medeiros, M. S., & A., S. (2018). Imagen del destino Natal, Brasil. Un análisis a partir del contenido generado por los usuarios en Tripadvisor. *Estudios y Perspectivas en Turismo, 27*, 533-549. Obtenido de https://www.redalyc.org/articulo.oa?id=180757123003
Mohammed, H., Rania, H., & Yasser, M. (2018). Sentiment Analysis of Social Media Networks Using Machine Learning. *International Computer Engineering Conference (ICENCO)*, 1-3. doi:10.1109/ICENCO.2018.8636124
Molina, M., Martínez, E., Martín, M., & Jiménez, S. (2015). eSOLHotel: Generación de un lexicón de opinión en español adaptado al dominio turístico. *Procesamiento del Lenguaje Natural*, 21-28. Obtenido de http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/5090/2971
Mridula, A., & Kavitha, C. (2018). Opinion Mining and Sentiment Study of Tweets Polarity Using Machine Learning. *Second International Conference on Inventive Communication and Computational Technologies (ICICCT)*. doi:10.1109/ICICCT.2018.8473079
Nilashi, M., Ibrahim, O., Yadegaridehkordi, E., Samad, S., Akbari, E., & Alizadeh, A. (2018). Travelers decision making using online review in social network sites: A case on TripAdvisor. *Journal of Computational Science*, 356-363. doi:10.1016/j.jocs.2018.09.006
Prerna, M., Ranjana, R., & Pankaj, K. (2019). Evaluating Performance of Machine Leaming Techniques used in Opinion Mining. *IEEE*. doi:10.1109/CCAA.2018.8777724
Ramanathan, V., & Meyyappan, T. (2019). Twitter Text Mining for Sentiment Analysis on People's Feedback about Oman Tourism. *IEEE*.

doi:10.1109/ICBDSC.2019.8645596
Samala, N., Shashanka, B., Shekhar, R., & Villamarin, R. (2020). Impact of AI and robotics in the tourism sector: a critical insight. *Journal of Tourism Futures*. doi:10.1108/JTF-07-2019-0065
Sánchez del Hoyo, R. (2019). Análisis de sentimientos con Twitter: turismo y política electoral. *Universidad de Sevilla*. Obtenido de https://idus.us.es/handle/11441/90023;jsessionid=90AE3476D921F8D764565AEA289DCCE7?
Shafqat-Ul-Ahsaan, Mourya, A., & Singh, P. (2019). Predictive Modeling and Sentiment Classification of Social Media Through Extreme Learning Machine. *Proceedings of ICETIT, 605*, 356-363. doi:10.1007/978-3-030-30577-2_30
Suganya, E., & Vijayarani, S. (2019). Sentiment Analysis for Scraping of Product Reviews from Multiple Web Pages Using Machine Learning Algorithms. *Intelligent Systems Design and Applications*. doi:10.1007/978-3-030-16660-1_66
Thelwall, M. (2019). Sentiment Analysis for Tourism. *Big Data and Innovation in Tourism, Travel, and Hospitality*, 87-104. doi:10.1007/978-981-13-6339-9_6
Topal, I., & Uçar, M. (2019). Hybrid Artificial Intelligence Based Automatic Determination of Travel Preferences of Chinese Tourists. *IEEE Access, 4*, 1-19. doi:10.1109/ACCESS.2019.2947712
Yang, X. (2019). Satisfaction Evaluation and Optimization of Tourism E-commerce Users Based on Artificial Intelligence Technology. *International Conference on Robots & Intelligent System (ICRIS)*, 373-375. doi:10.1109/ICRIS.2019.00100

6.2. GLOSARIOS DE TÉRMINOS

- Big Data: Grandes cantidades de datos, que posee características únicas, las cuales son volumen (grandes cantidades de datos), velocidad (la rapidez del procesamiento de esta información) y variedad (las diferentes características que poseen los datos encontrados).
- Ciencia de Datos: Aplicación de técnicas de programación para analizar datos.
- Minería de Datos: Técnicas utilizadas para encontrar la información más valiosa e importante de un gran conjunto de datos.
- Minería de Texto: Utiliza patrones observables en datos estructurados y almacenados en bases de datos relacionales. Información que no está explícita dentro del texto.
- Inteligencia Artificial: Es la tecnología que aplican las computadoras para simular las actividades humanas por cuenta propia, a través de un proceso de aprendizaje y entrenamiento.
- NLP: es la transformación e interpretación del lenguaje humano utilizando sistemas informáticos. Abarca parte de la Ciencia de Datos, Inteligencia Artificial (Aprendizaje Automático) y la lingüística.
- Análisis de Sentimiento: Es una técnica de machine learning, basada en el procesado del lenguaje natural, que pretende obtener información subjetiva de una serie de textos o documentos.
- Teorema de Bayes: Se basa en la probabilidad de un hecho ocurrido con anterioridad aplicando esta información para predecir algo que puede suceder.
- Deep Learning: Es un conjunto de aprendizajes que se basa en la arquitectura de redes neuronales.
- CNN: Es un tipo modelo de aprendizaje automático supervisado (machine learning) en el que el sistema es capaz de entender e interpretar tareas de clasificación ya sean, de textos, videos o imágenes.

- Clustering: Implica el agrupamiento de datos, a partir de esto se puede lograr la clasificación y también es una técnica para analizar grandes cantidades de información, además de ser un tipo de aprendizaje automático no-supervisado.
- K-Means: Es aprendizaje no supervisado que agrupa elementos k grupos basadas en las características en común.