



UNIVERSIDAD TÉCNICA PARTICULAR DE LOJA

La Universidad Católica de Loja

FACULTAD DE INGENIERÍAS Y ARQUITECTURA

CARRERA DE TECNOLOGÍAS DE LA INFORMACIÓN

**Clasificación de publicaciones en redes sociales mediante
el uso de técnicas de aprendizaje automático**

Trabajo de integración curricular previo a la obtención del título de:

INGENIERO EN TECNOLOGÍAS DE LA INFORMACIÓN

Autor: Ramos Argoti, Andrés Fernando

Director: Riofrio Calderón, Guido Eduardo

MANTA

2024



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

2024

Aprobación del director del Trabajo de Integración Curricular

Loja, 17 de Octubre de 2024

Doctor,

Daniel Alejandro Guamán Coronel

Director de la carrera de Informática y de Tecnologías de la Información

Ciudad. -

De mi consideración:

Me permito comunicar que, en calidad de director del presente Trabajo de Integración Curricular denominado: Clasificación de publicaciones en redes sociales mediante el uso de técnicas de aprendizaje automático realizado por Andrés Fernando Ramos Argoti ha sido orientado y revisado durante su ejecución, así mismo ha sido verificado a través de la herramienta de similitud académica institucional, y cuenta con un porcentaje de coincidencia aceptable. En virtud de ello, y por considerar que el mismo cumple con todos los parámetros establecidos por la Universidad, doy mi aprobación a fin de continuar con el proceso académico correspondiente.

Particular que comunico para los fines pertinentes.

Atentamente,

Director: Riofrio Calderón, Guido Eduardo, Magíster

C.I.: 1103214969

Correo electrónico: gerifrio@utpl.edu.ec

Declaración de autoría y cesión de derechos

Yo, Andrés Fernando Ramos Argoti, declaro y acepto en forma expresa lo siguiente:

Ser autor del Trabajo de Integración Curricular denominado: Clasificación de publicaciones en redes sociales mediante el uso de técnicas de aprendizaje automático, de la carrera de Tecnologías de la Información, específicamente de los contenidos comprendidos en: Capítulo uno: Antecedentes / generalidades / introducción, capítulo dos: marco teórico, capítulo tres: Metodología, capítulo cuatro: Desarrollo del proyecto y capítulo cinco: Análisis de resultados, siendo Guido Eduardo Riofrio Calderón, director del presente trabajo; también declaro que la presente investigación no vulnera derechos de terceros ni utiliza fraudulentamente obras preexistentes. Además, ratifico que las ideas, criterios, opiniones, procedimientos y resultados vertidos en el presente trabajo investigativo, son de mi exclusiva responsabilidad. Eximo expresamente a la Universidad Técnica Particular de Loja y a sus representantes legales de posibles reclamos o acciones judiciales o administrativas, en relación a la propiedad intelectual de este trabajo.

Que la presente obra, producto de mis actividades académicas y de investigación, forma parte del patrimonio de la Universidad Técnica Particular de Loja, de conformidad con el artículo 20, literal j), de la Ley Orgánica de Educación Superior; y, artículo 91 del Estatuto Orgánico de la UTPL, que establece: "Forman parte del patrimonio de la Universidad la propiedad intelectual de investigaciones, trabajos científicos o técnicos y tesis de grado que se realicen a través, o con el apoyo financiero, académico o institucional (operativo) de la Universidad", en tal virtud, cedo a favor de la Universidad Técnica Particular de Loja la titularidad de los derechos patrimoniales que me corresponden en calidad de autor/a, de forma incondicional, completa, exclusiva y por todo el tiempo de su vigencia.

La Universidad Técnica Particular de Loja queda facultada para ingresar el presente trabajo al Sistema Nacional de Información de la Educación Superior del Ecuador para su difusión pública, en cumplimiento del artículo 144 de la Ley Orgánica de Educación Superior.

.....

Autor: Andrés Fernando Ramos Argoti

C.I.: 0201347424

Correo electrónico: andres13ramos@hotmail.es

Dedicatoria

El presente trabajo lo dedico a la mujer que me dio la vida, ya que siempre tuvo la confianza y amor que tanto necesitaba para cumplir con esta meta, sin su cariño y oraciones esto no hubiera podido ser posible. Este gran logro va por ti querida madre como ofrenda por el buen trabajo que has hecho en guiarme día a día.

Agradecimiento

Primeramente, quiero dar un enorme agradecimiento a la universidad Técnica Particular de Loja ya que dentro de su alcance posee la formación a distancia con gran variedad de carreras para todos los interesados, esto hace que personas, así como en mi caso, que al tener que cumplir un horario laboral puedan cumplir la meta de formarse como profesionales.

También quiero agradecer a los docentes de cada asignatura, que a pesar de la limitante de no tener a los estudiantes presencialmente, han logrado llegar con su conocimiento utilizando las metodologías apropiadas.

Finalmente, agradezco a mis familiares y amigos que estuvieron presentes en este arduo camino y que muchas veces me dieron energía para continuar con una frase de apoyo y solidaridad.

Índice de contenido

Carátula.....	I
Aprobación del director del Trabajo de Integración Curricular.....	II
Declaración de autoría y cesión de derechos	III
Dedicatoria	IV
Agradecimiento.....	V
Índice de contenido	VI
Resumen.....	1
Abstract.....	2
Introducción.....	3
Capítulo uno.....	5
Introducción.....	5
1.1 Problemática	5
1.2 Objetivos	5
1.2.1 <i>Objetivo general</i>	5
1.2.2 <i>Objetivos específicos</i>	5
1.3 Componentes	6
1.3.1 <i>Componente de investigación</i>	6
1.3.2 <i>Componente de innovación</i>	6
1.3.3 <i>Componente técnico</i>	6
1.4 Estrategias o metodologías para el desarrollo	6
1.5 Estructura del documento.....	6
Capítulo dos.....	8
Marco teórico	8
2.1 Fundamentos de las redes sociales.....	8

2.1.1 Evolución de las redes sociales.....	9
2.1.2 Impacto social de las redes sociales	11
2.2 Aprendizaje automático	13
2.2.1 Tipos de aprendizaje automático	14
2.2.2 Técnicas de aprendizaje automático aplicadas a la clasificación	19
2.2.3 Modelos de aprendizaje automático para texto	20
2.3 Análisis de sentimientos	23
2.3.1 Métodos de análisis de sentimientos.....	24
2.3.2 Aplicaciones del análisis de sentimientos en redes sociales.....	27
2.4 Herramientas y tecnologías	29
2.4.1 Python y google colab.....	30
2.4.2 React js.....	30
2.4.3 Librerías y frameworks para aprendizaje automático.....	32
2.5 Trabajos relacionados.....	35
2.5.1 Estudios sobre clasificación de textos en redes sociales.....	35
2.5.2 Avances recientes en aprendizaje automático aplicado a redes sociales.....	38
Capítulo tres.....	41
Metodología	41
3.1 Diseño de la investigación	41
3.2 Recopilación de datos.....	42
3.3 Preprocesamiento de datos	43
3.4 Modelos de clasificación.....	44
Capítulo cuatro	45
Desarrollo del proyecto	45
4.1 Desarrollo del modelo de clasificación.....	45
4.2 Evaluación del modelo.....	45

4.3 Visualización de resultados	49
4.4 Pruebas del sistema.....	59
Capítulo cinco.....	61
Análisis de resultados	61
5.1 Discusión de resultados	61
5.2 Comparación con trabajos anteriores	63
Conclusiones	65
Recomendaciones	66
Referencias	67

Índice de tablas

Tabla 1 Tipos de aprendizaje automatizado	15
Tabla 2 Métodos de análisis de sentimientos.....	24
Tabla 3 Aplicación de análisis de sentimientos	28
Tabla 4 Ejemplos de librerías y frameworks.....	32
Tabla 5 Tabla resumen de los modelos	60
Tabla 6 Tendencia del modelo Naive Bayes (MNB)	61
Tabla 7 Tendencia del modelo Naive Bayes (MNB)	61
Tabla 8 Tendencia del modelo Support Vector Machines (SVM).....	62
Tabla 9 Tendencia del modelo Support Vector Machines (SVM).....	62
Tabla 10 Tendencia del modelo Random Forest (RF).....	62
Tabla 11 Tendencia del modelo Random Forest (RF).....	63

Índice de figuras

Figura 1 Imagen de la evolución de las redes sociales	11
Figura 2 Ejemplo de técnica de regresión	16
Figura 3 Imagen de un modelo de detector de spam	17
Figura 4 Imagen de un modelo de aprendizaje semisupervisado	18
Figura 5 Modelos de aprendizaje automáticos	21
Figura 6 Modelos de aprendizaje automático	22
Figura 7 Imagen de React JS	31
Figura 8 Imagen del Proceso KDD	41
Figura 9 Imagen de ejemplo de datasets	42
Figura 10 Naive Bayes Resultados con el 5% dataset.....	46
Figura 11 Naive Bayes Resultados con el 20% dataset.....	47
Figura 12 SVM Resultados con el 5% dataset	47
Figura 13 SVM Resultados con el 20% dataset	48
Figura 14 Random Forest Resultados con el 5% dataset.....	48
Figura 15 Random Forest Resultados con el 20% dataset.....	49
Figura 16 Matriz de confusión – MNB con el 5% del dataset.....	50
Figura 17 Matriz de confusión – MNB con el 20% del dataset.....	51
Figura 18 Curva ROC - MNB	51
Figura 19 Curva de precisión sensibilidad – MNB.....	52
Figura 20 Matriz de confusión – SVM con el 5% del dataset.....	53
Figura 21 Matriz de confusión – SVM con el 20% del dataset	53
Figura 22 Curva ROC - SVM.....	54
Figura 23 Curva de precisión sensibilidad – SVM.....	55
Figura 24 Matriz de confusión – RF con el 5% del dataset.....	56
Figura 25 Matriz de confusión – RF con el 20% del dataset.....	57

Figura 26 Curva ROC - RF..... 58
Figura 27 Curva de precisión sensibilidad – RF..... 59

Resumen

Por medio del aprendizaje automatizado se puede establecer clasificaciones y predicciones de alto nivel, sin embargo, esta herramienta aún no ha sido utilizada en las diversas plataformas digitales como las redes sociales que permitan clasificar la información. Como objetivo del presente proyecto fue clasificar publicaciones en redes sociales mediante el uso de técnicas de aprendizaje automático. La metodología fue aplicada y descriptiva para el análisis de los sentimientos en las publicaciones por medio de los modelos Bayes, Support Vector Machine (SVM) y Random Forest. Como resultados se obtuvo que los tres modelos brindan una exactitud sobre el 74% lo que señala una alta capacidad exactitud en la clasificación de las publicaciones en redes sociales. Se concluye que el modelo que entrega mejores resultados con los datos de este trabajo es el de vectores o SVM. No obstante, en el tiempo de predicción se ve que el más rápido es el modelo de redes bayesianas.

Palabras clave: Aprendizaje automatizado, redes sociales, Modelos.

Abstract

Through machine learning, high-level classifications and predictions can be established. However, this tool has not yet been widely used on various digital platforms, such as social media, for information classification. The objective of this project was to classify social media posts using machine learning techniques. The methodology was applied and descriptive, analyzing sentiments in posts through Bayesian models, Support Vector Machine (SVM), and Random Forest. Results showed that all three models provided accuracy above 0.74, indicating a high classification capacity. The SVM model yielded the best results with the given data, while the Bayesian network model was the fastest in prediction time.

Keywords: Machine learning, social media, models.

Introducción

Un cambio trascendental que se ha visto en la sociedad es el uso de las redes sociales, este es el medio de comunicación predilecto en la actualidad, la variedad de formas de interacción acoge a personas de todas las edades, así como: mensajes de texto, mensajes de audio, llamadas de audio o video llamadas.

Para abordar el problema de investigación, el proyecto se centró en la clasificación de publicaciones en redes sociales utilizando técnicas de aprendizaje automático. A través de una profunda investigación, se han aplicado modelos de aprendizaje automático para analizar datos y categorizar publicaciones. Los resultados obtenidos pueden ser utilizados en varios contextos de manera innovadora, beneficiando así a los usuarios de redes sociales.

El proyecto logró cumplir con todos los objetivos planteados inicialmente. La revisión sistemática de técnicas de aprendizaje automático proporcionó una base sólida para seleccionar los métodos más adecuados. La recopilación de datos mediante modelos de aprendizaje automático se llevó a cabo de manera eficiente, permitiendo un análisis profundo de las publicaciones en redes sociales. Finalmente, la validación de los resultados mediante métricas de evaluación demostró el correcto funcionamiento de los modelos de clasificación, destacando especialmente el rendimiento del modelo SVM.

El desarrollo del proyecto se benefició de la disponibilidad de herramientas avanzadas y bibliotecas de aprendizaje automático que facilitaron la implementación y comparación de diversos modelos. Además, el acceso a grandes volúmenes de datos de redes sociales proporcionó una amplia base de datos para entrenar y probar los modelos. Estas facilidades permitieron explorar diferentes enfoques y optimizar los resultados de manera efectiva.

Sin embargo, el proyecto enfrentó varios desafíos. La variabilidad y el ruido en los datos de redes sociales complicaron el proceso de preprocesamiento y limpieza de los datos. También fue necesario ajustar y optimizar constantemente los modelos para mejorar la precisión y eficiencia, lo cual requirió un esfuerzo considerable en términos de tiempo y recursos. A pesar de estas limitaciones, el proyecto logró alcanzar sus objetivos, demostrando

la viabilidad y eficacia del uso de técnicas de aprendizaje automático para la clasificación de publicaciones en redes sociales.

La metodología utilizada fue aplicada y descriptiva lo cual permitió aplicar los modelos de aprendizaje automatizado en bases de datos de redes sociales para identificar si estos sistemas permiten la clasificación de las publicaciones, lo cual sus resultados fueron desglosados detalladamente para su análisis y así obtener las conclusiones y recomendaciones del proyecto.

Para el desarrollo de este proyecto de investigación se desarrolló en cinco capítulos, lo cual el capítulo uno corresponde a la identificación del problema de investigación y los objetivos; en el capítulo dos se desarrolló el marco teórico que fueron las bases para establecer los modelos a aplicarse en el aprendizaje automatizado, en el capítulo tres se desarrolló la metodología que se requirió en el proyecto; en el capítulo cuatro se realizó el desarrollo de la aplicación del modelo de aprendizaje automatizado para la clasificación, finalmente en el capítulo cinco se expusieron los principales resultados que se observó en los modelos, y así describir las conclusiones y recomendaciones de la investigación.

La investigación sobre la clasificación de publicaciones en redes sociales mediante técnicas de aprendizaje automático es crucial para las empresas, ya que optimiza la gestión de información y estrategias de marketing. Para los usuarios, mejora la experiencia en línea al filtrar contenido irrelevante y evitar la desinformación. A nivel de sociedad, esta investigación facilita la gestión de crisis al permitir la rápida identificación de información relevante y contribuye a un mejor discurso público al priorizar contenido fiable. En general, ofrece beneficios significativos para la eficiencia y calidad de la información en redes sociales.

Capítulo uno

Introducción

1.1 Problemática

No cabe duda de que el uso de las redes sociales en los últimos años ha tenido un alto crecimiento, es claro que con la expansión de las tecnologías llegará al alcance de todos. Existen innumerables ventajas del uso de las redes sociales siendo la comunicación el pilar fundamental, también es una herramienta importante para el marketing tanto para publicar artículos de venta como para buscar algo que se desea comprar, de la misma manera se pueden enumerar algunas desventajas como peligros a la seguridad, riesgos a la imagen personal, críticas, falsa información, entre otras. Para los usuarios de las redes sociales es una molestia estar infestados con publicaciones que no son de interés o que son falsos y que resultan una pérdida de tiempo para el lector.

Sería de gran beneficio para los usuarios de redes sociales poder categorizar por importancia o simplemente desechar las publicaciones que no son de interés. Mediante una profunda investigación del aprendizaje automático se aplicarían técnicas para el análisis de datos que se puede incorporar a las redes sociales, estos resultados obtenidos se pueden utilizar en varios contextos de una manera innovadora para beneficio del usuario.

1.2 Objetivos

1.2.1 *Objetivo general*

Clasificar publicaciones en redes sociales mediante el uso de técnicas de aprendizaje automático.

1.2.2 *Objetivos específicos*

- Realizar una investigación sobre técnicas de aprendizaje automático, mediante una revisión sistemática de información.
- Realizar la obtención de datos a través de un modelo de aprendizaje automático.

- Analizar publicaciones en redes sociales para su posterior clasificación.
- Validar los resultados del modelo de clasificación mediante métricas de evaluación para establecer el correcto funcionamiento.

1.3 Componentes

1.3.1 Componente de investigación

Se realizará investigación en técnicas de aprendizaje automático, de esta manera hacer uso de la predicción y clasificación para detectar patrones en los datos obtenidos en los comentarios y reacciones de las publicaciones en las redes sociales.

1.3.2 Componente de innovación

Es innovador aportar con herramientas para realizar una clasificación de las publicaciones en cuentas de redes sociales, una vez obtenida una clasificación existe variedad de aplicaciones que se pueden dar en base a este tema.

1.3.3 Componente técnico

Para el desarrollo de este proyecto se utilizará técnicas de aprendizaje automático a través del lenguaje Python en el entorno Google Colab

1.4 Estrategias o metodologías para el desarrollo

Se decide utilizar un enfoque cuantitativo para este tema de TIC, ya que se encuentra originada en una idea que con el tiempo se ha ido delimitando, además se realizará una revisión literaria para construir un marco teórico sustentable, por otro lado, los resultados obtenidos de este proyecto serán medibles y cuantificables, finalmente el proceso de desarrollo será a través de un orden secuencial y probatorio.

1.5 Estructura del documento

Este proyecto se desarrolló en cinco capítulos, en donde el capítulo uno corresponde a la identificación del problema de investigación y los objetivos; en el capítulo dos se desarrolla el marco teórico, en el capítulo tres se desarrolla la metodología que se requirió en el proyecto; en el capítulo cuatro se realiza el desarrollo de la aplicación del modelo de

aprendizaje automático, finalmente en el capítulo cinco se expusieron los principales resultados que se observó en los modelos.

Capítulo dos

Marco teórico

2.1 Fundamentos de las redes sociales

En los últimos años la percepción de red social ha cobrado una gran importancia, por lo que ahora es una expresión común que asociamos con plataformas como Whatsapp, Facebook o Twitter con la actualidad, no obstante, su presencia va más allá de la intercomunicación y es más compleja. Las redes sociales han sido objeto de estudio en diversas disciplinas a lo largo del tiempo, por lo que se han desarrollado teorías para explicar su funcionamiento en la base a su evolución en el entorno virtual. Además, con la llegada de la Web 2.0, las redes sociales en línea se han vuelto fundamentales en las relaciones interpersonales representando la posibilidad de utilizar e interpretar Internet de una nueva forma (Marín y Cabero, 2019).

Las Tecnologías de la Información y Comunicación (TIC) han evolucionado significativamente, dando lugar a la aparición de la Web 2.0, que se caracteriza por ser un conjunto de herramientas que facilitan la creación colaborativa de conocimiento, integrando tres elementos clave: tecnología, conocimiento y usuarios, generando espacios de conexión propicios para el desarrollo de una nueva forma de alfabetización tecnológica que fomenta la crítica, la colaboración y la creatividad, y a su vez promueve la creación colectiva de contenidos, el uso de recursos compartidos, además del control de calidad entre los usuarios. En este sentido, las herramientas de la Web 2.0 permiten tanto la interacción como la asistencia entre usuarios, contribuyendo al aprendizaje e intercomunicación, estos canales incluyen blogs, diarios personales, canales de transmisión (Facebook, Instagram, X, Tiktok, Whatsap) y wikis (Rodríguez et al., 2019).

La particularidad de las redes sociales radica en la creación de comunidades virtuales donde los individuos pueden establecer vínculos, ya sea con amigos, familiares, colegas o incluso desconocidos con intereses similares, lo que ha dado lugar a la formación de redes

de apoyo, colaboración o participación, abriendo puertas a nuevas oportunidades informativas, de aprendizaje, desarrollo personal o profesional. Además, estas desempeñan un papel fundamental en la difusión de noticias, novedades, promoción de eventos, comercialización de productos o servicios, y la generación de intercambios culturales significativas que pueden influir en la opinión pública o hasta global (Tabassum et al., 2018).

En un amplio sentido, las redes sociales son una estructura social formada tanto por personas como por entidades interconectadas que se encuentran entre la convergencia de la tecnología digital y la necesidad humana de conectarse, basado en esto, estas plataformas han cambiado la manera en que interactuamos, brindando a los individuos un espacio que puedan compartir no solo información personal, sino también ideas, conocimientos al igual que opiniones sobre una amplia gama de temas.

2.1.1 Evolución de las redes sociales

Las redes sociales han experimentado una evolución notable desde su surgimiento como una herramienta de comunicación emergente hasta convertirse en una parte indispensable de la vida diaria de los usuarios. Inicialmente, surgieron como parte del marketing 2.0, marcando una transición de la web estática y unidireccional de la web 1.0 a la web 2.0, donde se introdujeron las interacciones personales y la humanización de las marcas a través de las redes sociales, esta evolución ha permitido que las redes sociales se conviertan en espacios dinámicos donde los usuarios pueden conectar, compartir contenido, intereses y disfrutar de formas variadas de entretenimiento. Además, durante su evolución las redes sociales han transformado la comunicación y la estrategia de interacción de diversos campos, brindando la oportunidad de conectar tanto directa como rápidamente, generando contenido especializado y aprovechando herramientas de análisis (Pérez et al., 2019).

La evolución de las redes sociales ha sido marcada por varios hitos importantes a lo largo de los años. En sus inicios, estas plataformas se centraban principalmente en la interacción social entre usuarios, ofreciendo herramientas básicas como la publicación de

mensajes o fotos, sin embargo, con el tiempo han evolucionado hacia plataformas más complejas y multifacéticas. La integración de las redes sociales en el marketing empresarial ha permitido aprovechar la conectividad web y la interactividad con los usuarios con propósitos comerciales. Esto ha resultado en un aumento significativo de compañías que utilizan estos canales de comunicación, motivados por el contacto directo con los consumidores y la efectividad en la estimulación de actividades comerciales. Plataformas como Facebook y Twitter se destacan como las más utilizadas por las empresas, lo que ha impulsado la importancia del Marketing de Redes Sociales (MRS) como parte fundamental de las estrategias comerciales. Esta evolución ha generado un interés creciente en la comunidad científica, que busca comprender y aprovechar el uso del MRS dentro de las organizaciones en un mercado altamente competitivo (Gómez y Palacios, 2021).

La evolución de las redes sociales ha sido un fenómeno complejo y dinámico que ha transformado la forma en que las personas se relacionan, se informan e incluso en como participan en la sociedad digital, ya que, en sus inicios las redes sociales surgieron como plataformas simples para la interacción social en línea, permitiendo a los usuarios conectarse, compartir mensajes, fotos o hasta establecer relaciones virtuales, sin embargo, con el tiempo estas plataformas evolucionaron hacia entornos más interactivos y ricos en contenido, incorporando características - herramientas nuevas de edición avanzada para el contenido que se comparte en este medio. Esta evolución no solo se ha centrado en la experiencia del usuario, sino también en la diversificación de las redes sociales a fin de atender diferentes necesidades de los diferentes tipos de audiencias. De esta forma, surgieron redes especializadas como LinkedIn, orientada a profesionales además de networking laboral; Instagram, enfocada en la creatividad visual o la expresión artística; Twitter, direccionada a la difusión de información en tiempo real permitiendo debate público; y plataformas como TikTok, que revolucionaron la forma de consumir contenido a través de videos cortos que se hacen virales (Moreno, 2019).

Figura 1

Imagen de la evolución de las redes sociales



Nota Elana BS. (2018). Evolución de las redes sociales [Fotografía]. Unsplash.

<https://www.istockphoto.com/es/vector/las-personas-de-negocios-y-tecnolog%C3%ADa-gm953443738-260266774>

La evolución de las redes sociales ha sido impresionante, desde sus inicios como simples plataformas de conexión hasta convertirse en poderosas herramientas de comunicación, interacción y marketing empresarial, estas diversas plataformas han transformado radicalmente la forma en que nos comunicamos, relacionamos, compartimos información y consumimos contenido, impactando significativamente la sociedad, la cultura y la economía a nivel global.

2.1.2 Impacto social de las redes sociales

Las redes sociales han impactado profundamente en la sociedad contemporánea al facilitar la comunicación instantánea y el intercambio de información, así como al fomentar la creación de comunidades virtuales. Este fenómeno ha llevado a un rápido crecimiento y una presencia omnipresente en la vida de las personas. Sin embargo, también se ha planteado que las redes sociales pueden ser un reflejo de una sociedad insatisfecha, donde la búsqueda de felicidad a través del consumo y la apariencia es prevalente. Este doble papel de las redes

sociales como herramientas de conexión y como potenciales generadores de vacío emocional y superficialidad plantea desafíos significativos en cuanto a su impacto en la salud mental, la privacidad y la formación de identidades individuales y colectivas. Es necesario reflexionar sobre este fenómeno para evitar efectos negativos y promover un uso más consciente y positivo de las redes sociales en beneficio de la sociedad en su conjunto (Blasco, 2021).

El impacto de las redes sociales en la generación Z ha sido objeto de estudio, destacando la preocupación por el potencial adictivo de estas plataformas. Los psicólogos y expertos tecnológicos han alertado sobre los riesgos de la adicción, evidenciando que el uso excesivo puede conducir a ansiedad, dependencia emocional, pérdida de motivación, falta de autocontrol e irritabilidad. Este fenómeno se relaciona con la capacidad adictiva de las redes sociales, incluso mayor que la del tabaco según estudios como el realizado por la Chicago Booth School of Business en 2012. La búsqueda de validación a través de likes y comentarios se ha convertido en una parte significativa de la experiencia en redes sociales para esta generación, generando impactos en su autoestima, comportamiento y salud mental. Es crucial continuar investigando y concienciando sobre estos efectos para fomentar un uso responsable y saludable de las redes sociales entre los jóvenes (Critikián y Medina, 2021).

El impacto social de las redes sociales es profundo y multifacético pues, por un lado, han transformado tanto la forma en que nos comunicamos como la manera en la que compartimos información, facilitando la conectividad a nivel global generando gran impacto en diferentes campos de interés, esto ha tenido un impacto positivo al permitir que las personas se mantengan en contacto, compartan experiencias e incluso encuentren apoyo, no obstante, también han surgido desafíos, como la propagación de información falsa, el aumento de la ansiedad o la dependencia de algunos usuarios a las redes, y la pérdida de privacidad evidenciando el aumento en casos de acoso, secuestro y agresión.

2.2 Aprendizaje automático

El aprendizaje automático (AA) es una rama fundamental de la inteligencia artificial (IA) que ha experimentado un rápido desarrollo en las últimas décadas, este surgió como una respuesta a la necesidad de que las máquinas puedan aprender e incluso mejorar su desempeño sin una programación explícita. Esta capacidad de aprendizaje ha sido definida como el medio que permite a los computadores aprender sin ser programados específicamente para cada tarea, o también como un programa de computador que perfecciona su rendimiento en una tarea específica a medida que adquiere experiencia en ella. El AA ha crecido de manera significativa gracias a la abundancia de datos disponibles en la era del Big Data y el acceso a Internet y se ha dividido en varias categorías para abordar diferentes tipos de problemas como; el aprendizaje supervisado, donde se entrena a los modelos con datos etiquetados y se les enseña a hacer predicciones basadas en ejemplos previos; el ANS, que se centra en identificar patrones o estructuras en los datos sin la guía de etiquetas predefinidas; y el aprendizaje por refuerzo, que utiliza la retroalimentación del entorno para mejorar tanto las decisiones como acciones de un agente (Díaz, 2021).

Uno de los desarrollos más significativos en el AA es el aprendizaje profundo (AP), que se basa en redes neuronales artificiales (RNA) con múltiples capas ocultas, que han demostrado una capacidad impresionante para resolver problemas complejos en áreas como la visión computacional, el procesamiento de lenguaje natural y la interpretación de datos. Las redes neuronales artificiales del aprendizaje profundo más utilizadas incluyen el Perceptrón Multicapa, las Redes Neuronales Convolucionales (CovNet), las Redes Neuronales Recurrentes y las Redes Generativas Antagónicas, asimismo, el desarrollo del AA y el AP ha sido impulsado por avances en hardware, como las unidades de procesamiento gráfico (GPU) que permiten un procesamiento más rápido y eficiente de datos. Además, plataformas en línea como Google Colab y Azure Microsoft han facilitado el acceso e implementación de algoritmos de AA y AP (Álvarez et al., 2020).

El aprendizaje automático, también conocido como machine learning, es una disciplina que ha demostrado ser una realidad en diversas áreas, como finanzas, mercadotecnia y ciencias sociales, donde ha generado resultados prometedores al predecir comportamientos. Sin embargo, su aplicación en medicina enfrenta desafíos significativos debido a problemas en el desarrollo de modelos, así como a cuestiones legales, éticas y filosóficas. A pesar de estos obstáculos, la complejidad inherente a la medicina la convierte en uno de los campos que más puede beneficiarse de estas técnicas. En los últimos años, se han incorporado diferentes modelos y algoritmos de aprendizaje automático en procesos médicos, como la detección de cáncer de mama mediante mamografías, el análisis de imágenes para la detección de cáncer de piel, y la evaluación de ecocardiogramas para la detección de patologías cardíacas. Esta integración creciente de técnicas de aprendizaje automático en medicina plantea desafíos importantes, pero también promete avances significativos en el diagnóstico y tratamiento de enfermedades (Quiró et al., 2020)

El aprendizaje automático es una rama de la inteligencia artificial que se centra en el desarrollo tanto de algoritmos como de modelos computacionales, que son capaces de aprender e incluso mejorar su desempeño a partir de los datos, mediante el uso de técnicas estadístico computacionales que les permiten a las máquinas identificar patrones, tendencias y relaciones en los datos, sin requerir una programación explícita.

2.2.1 Tipos de aprendizaje automático

Según Álvarez (2021), el aprendizaje automático se puede clasificar en varios tipos según la forma en que se realiza el proceso de aprendizaje y la disponibilidad de datos etiquetados. A continuación, se detallan los principales tipos de aprendizaje automático:

Tabla 1

Tipos de Aprendizaje Automatizado

Tipo	Concepto
Aprendizaje Supervisado:	En este tipo de aprendizaje, el modelo se entrena utilizando un conjunto de datos etiquetados, es decir, cada ejemplo de entrenamiento tiene una etiqueta que indica la respuesta deseada. El objetivo del modelo es aprender a mapear las entradas a las salidas basándose en ejemplos proporcionados. Algunos algoritmos comunes de aprendizaje supervisado incluyen las redes neuronales, los árboles de decisión y las máquinas de vectores de soporte (SVM) (Álvarez et al., 2021).
Aprendizaje No Supervisado:	En contraste con el aprendizaje supervisado, en el aprendizaje no supervisado no se proporcionan etiquetas en los datos de entrenamiento. El modelo debe encontrar patrones y estructuras ocultas en los datos por sí mismo. Este tipo de aprendizaje es útil para la segmentación de datos, la detección de anomalías y la reducción de dimensionalidad. Algunos algoritmos no supervisados comunes son el clustering (agrupamiento) y la reducción de la dimensionalidad como PCA (Análisis de Componentes Principales) y t-SNE (t-Distributed Stochastic Neighbor Embedding) (Álvarez et al., 2021).
Aprendizaje por Refuerzo:	En el aprendizaje por refuerzo, el modelo aprende a través de la interacción con un entorno dinámico, pues recibe recompensas según las acciones que realice, lo que le permite aprender a tomar decisiones que maximicen la recompensa a largo plazo. Este tipo de aprendizaje se utiliza en aplicaciones como juegos, robótica y control de procesos. Algoritmos de aprendizaje por refuerzo incluyen Q-Learning, Deep Q-Networks (DQN) y Policy Gradient methods (Álvarez et al., 2021).

Nota. Esta tabla muestra los principales tipos de aprendizaje automático.

Además de estos tipos principales, existen variantes y combinaciones de estos tipos, como el aprendizaje semi-supervisado (que utiliza datos etiquetados y no etiquetados), el

aprendizaje activo (donde el modelo puede seleccionar qué datos quiere etiquetar) y el aprendizaje transferido (que aprovecha el conocimiento aprendido en un dominio para otro relacionado). Estos diferentes tipos de aprendizaje automático tienen aplicaciones diversas y permiten abordar una amplia gama de problemas en campos como la visión por computadora, el procesamiento del lenguaje natural, la salud, la seguridad, entre otros (Álvarez et al., 2021).

Por su parte Pineda (2022), menciona que existen varias categorías de aprendizaje automático, siendo las más notables: supervisado, no supervisado y por refuerzo.

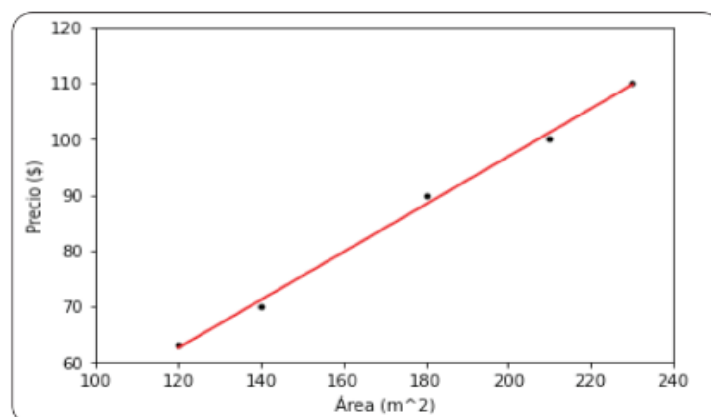
En el aprendizaje supervisado: cada muestra del conjunto de datos está etiquetada con un resultado específico o una categoría, utilizando una variable llamada destino u objetivo. Estos datos etiquetados permiten que el modelo aprenda patrones internos para posteriormente hacer predicciones específicas, dando resultados con bajo índice de error.

Dentro del aprendizaje supervisado, se pueden distinguir dos enfoques:

Regresión: Este enfoque permite predecir valores continuos considerando múltiples variables de entrada, buscando establecer una relación entre una variable dependiente y varias variables independientes, basado en esto algunos problemas se pueden abordar mediante técnicas de regresión (Pineda, 2022).

Figura 2

Ejemplo de técnica de regresión

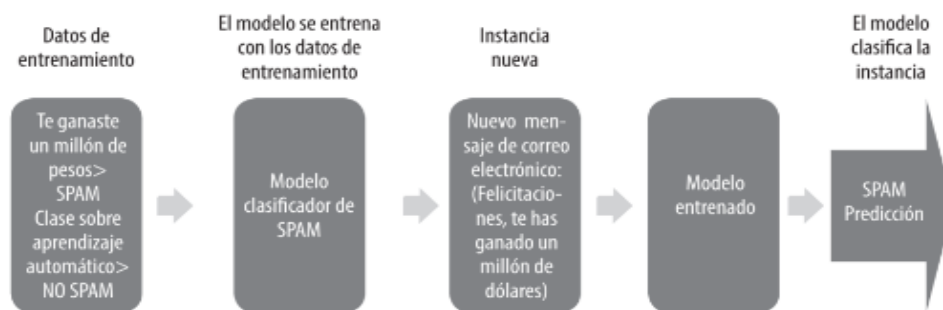


Nota. Adoptado de Aprendizaje automático y profundo en Python Una mirada hacia la inteligencia artificial [fotografía], por Pineda, 2022,

Clasificación: Es un tipo de algoritmo en el cual el dato a predecir es discreto, por lo que se le asocia a una categoría o clase a la cual pertenece, como un modelo de detector de Spam esquematizado (Pineda, 2022).

Figura 3

Imagen de un modelo de detector de spam



Nota. Adoptado de Aprendizaje automático y profundo en Python
Una mirada hacia la inteligencia artificial [fotografía], por Pineda,
2022, RA-MA Editorial

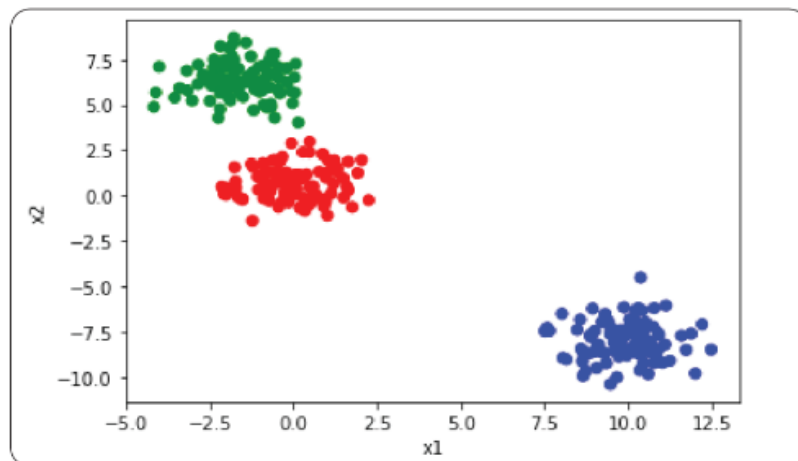
Aprendizaje no supervisado: Este modelo se caracteriza por la ausencia de etiquetas en las muestras de entrenamiento por lo que, trabaja directamente con los datos de entrada y busca identificar relaciones basadas en características comunes. Su principal objetivo es abordar dos problemas principales: el agrupamiento o clustering, que consiste en agrupar los datos según un criterio de distancia o similitud por medio de algoritmos destacados como el k-medias (k-means en inglés); y la reducción de la dimensionalidad, que implica la aplicación de algoritmos que comprimen el número de variables a fin de obtener un espacio multidimensional más condensado y lograr resultados más precisos (Pineda, 2022).

Este se aplica en diversos campos, como el diagnóstico médico y la segmentación de clientes en marketing para ofrecer productos o servicios personalizados. Otros algoritmos de agrupamiento, como DBSCAN, también son útiles en la detección de anomalías o más.

Aprendizaje semisupervisado: Se utiliza tanto en datos etiquetados como en no etiquetados, lo que en ciertas situaciones puede resultar en modelos más precisos (Pineda, 2022).

Figura 4

Imagen de un modelo de aprendizaje semisupervisado



Nota. Adoptado de Aprendizaje automático y profundo en Python Una mirada hacia la inteligencia artificial [fotografía], por Pineda, 2022, RA-MA Editorial.

Aprendizaje por refuerzo: Se basa en la experimentación y corrección de errores, donde cada vez que se logra una predicción acertada se otorga una recompensa, y en caso contrario, se aplica una penalización, enfoque ensayo - error. Los algoritmos de aprendizaje en este contexto, buscan desplegar estrategias a fin de minimizar las penalizaciones y maximizar las recompensas, con el propósito de mejorar su capacidad predictiva (Pineda, 2022).

El aprendizaje automático se centra en desarrollar algoritmos y técnicas que permiten a las computadoras aprender y mejorar su desempeño a partir de datos. Dentro del aprendizaje automático, existen diferentes tipos que se utilizan según la problemática y la naturaleza de los datos disponibles, estos incluyen; el aprendizaje supervisado, donde el

modelo se entrena con ejemplos etiquetados para predecir etiquetas de datos nuevos; el aprendizaje no supervisado, que busca patrones en datos no etiquetados para encontrar relaciones; y el aprendizaje por refuerzo, en el que un agente aprende a través de la interacción con un entorno, recibiendo recompensas o penalizaciones.

2.2.2 Técnicas de aprendizaje automático aplicadas a la clasificación

El aprendizaje automático, especialmente cuando se emplea en técnicas de clasificación, ha revolucionado la forma en que abordamos la organización - comprensión de datos complejos, ya que, permiten que los sistemas informáticos aprendan patrones y características distintivas de conjuntos de datos masivos, lo que les capacita para realizar clasificaciones precisas y automatizadas. Por ello, en el contexto de la clasificación de objetos reciclables, el aprendizaje automático permite entrenar modelos de redes neuronales convolucionales (CNN) que son capaces de distinguir entre diferentes tipos de materiales, como plástico, papel, vidrio y metal, incluso cuando presentan variaciones en forma, tamaño o color. Además, el uso de técnicas avanzadas como el aprendizaje por transferencia (Transfer Learning) permite aprovechar el conocimiento previamente adquirido acelerando el proceso de entrenamiento al mismo tiempo que mejora la precisión de las clasificaciones (Salamanca, 2021).

Estas aplicaciones se vuelven aún más poderosas cuando se combinan con herramientas y lenguajes de programación como Python, TensorFlow y Keras, proporcionando una plataforma robusta y escalable para implementar soluciones de clasificación de objetos reciclables en entornos prácticos, como el uso de Raspberry Pi y módulos de cámara picam para la detección en tiempo real en instalaciones de reciclaje o puntos de recolección de residuos. De esta manera, el aprendizaje automático aplicado a la clasificación no solo optimiza los procesos de reciclaje al facilitar la identificación y separación de materiales, sino que también contribuye significativamente a la promoción de prácticas

sostenibles y a la reducción del impacto ambiental al fomentar la reutilización y el reciclaje eficiente de recursos (Ruiz , 2019).

Las técnicas avanzadas de aprendizaje automático, destacando el uso de redes neuronales convolucionales (CNN) para la clasificación de imágenes de objetos reciclables, son efectivas en tareas de visión por computadora debido a su capacidad para aprender patrones complejos, ya que, fueron entrenados para realizar clasificaciones, distinguiendo entre objetos reciclables y no reciclables, identificando distintas categorías como plástico, vidrio, metal, papel-cartón, orgánico y no reciclable. Además de la implementación de CNN, se exploró el concepto de aprendizaje por transferencia (Transfer Learning), que permite aprovechar el conocimiento adquirido por modelos previamente entrenados en conjuntos de datos más grandes y generalizados, lo que permitió acelerar el proceso de entrenamiento a fin de mejorar la precisión de las clasificaciones. Por medio de programas como Python, aprovechando las capacidades de los frameworks de aprendizaje automático como TensorFlow y la librería de alto nivel Keras (Salina et al., 2021).

Las técnicas de aprendizaje automático son fundamentales en la era digital actual, ya que se centran en entrenar algoritmos para reconocer patrones y tomar decisiones en función de datos históricos. Entre las técnicas más utilizadas se encuentran la clasificación binaria y multiclase, donde se asignan etiquetas a datos según categorías predefinidas. Además, el modelado de redes neuronales convolucionales (CNN) ha demostrado eficacia en la extracción de características relevantes en imágenes para su clasificación.

2.2.3 Modelos de aprendizaje automático para texto

Según Kadhim (2019), Los modelos de aprendizaje automático para texto son herramientas poderosas que utilizan algoritmos para analizar y comprender datos textuales de manera automatizada. Algunos de los modelos más utilizados son:

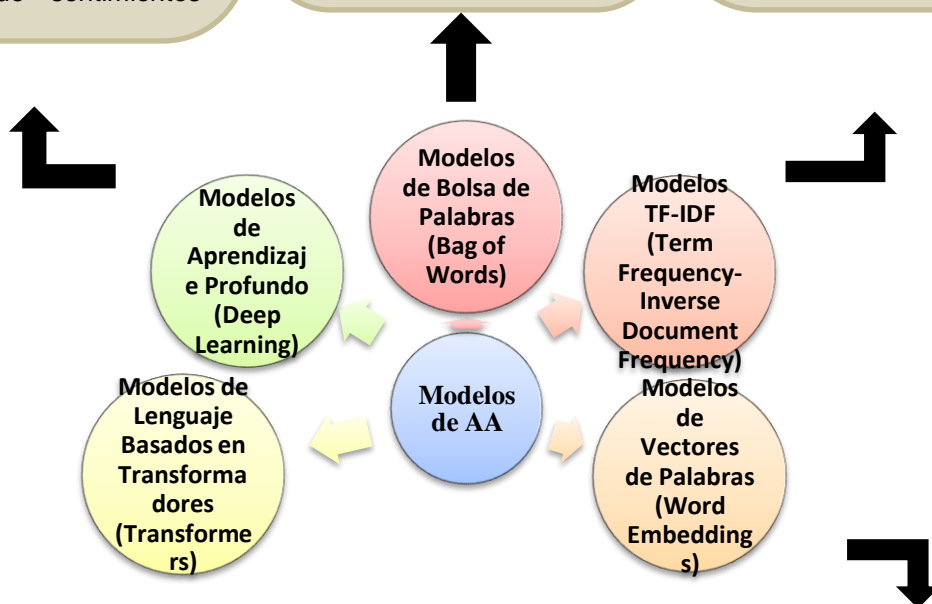
Figura 5

Modelos de Aprendizaje Automáticos

Las redes neuronales recurrentes (RNN), las redes neuronales convolucionales (CNN) y las redes neuronales recurrentes bidireccionales (Bi-RNN) son ejemplos de modelos de aprendizaje profundo utilizados para tareas de procesamiento de texto como la traducción automática, la generación de texto y el análisis de sentimientos

Este enfoque representa el texto como una colección de palabras sin considerar el orden o la estructura gramatical. Cada palabra se convierte en una característica y se cuenta su frecuencia en el documento (De Francisco, 2018).

Estos modelos asignan un peso a cada palabra en función de su frecuencia en el documento y en el corpus completo de documentos. Las palabras más frecuentes en el documento, pero poco comunes en el corpus obtienen un peso más alto (Zaid et al., 2022).



Los modelos como BERT utilizan arquitecturas basadas en transformadores para capturar el contexto y las relaciones semánticas en el texto. Estos modelos han demostrado un rendimiento excepcional en tareas como la comprensión del lenguaje natural y la generación de texto (Beltrán y Rodríguez, 2021).

Estos modelos representan las palabras como vectores densos en un espacio multidimensional donde palabras similares tienen vectores cercanos. Word2Vec, GloVe y FastText son ejemplos populares de este tipo de modelos

Nota. Esta es una recopilación de los modelos de aprendizaje automático para texto.

Cada modelo tiene sus ventajas y desventajas dependiendo del contexto y la tarea específica que se quiera abordar en el procesamiento de texto. La elección del modelo adecuado suele depender de factores como el tamaño del corpus, la complejidad del lenguaje y los objetivos del análisis textual (Kadhim, 2019).

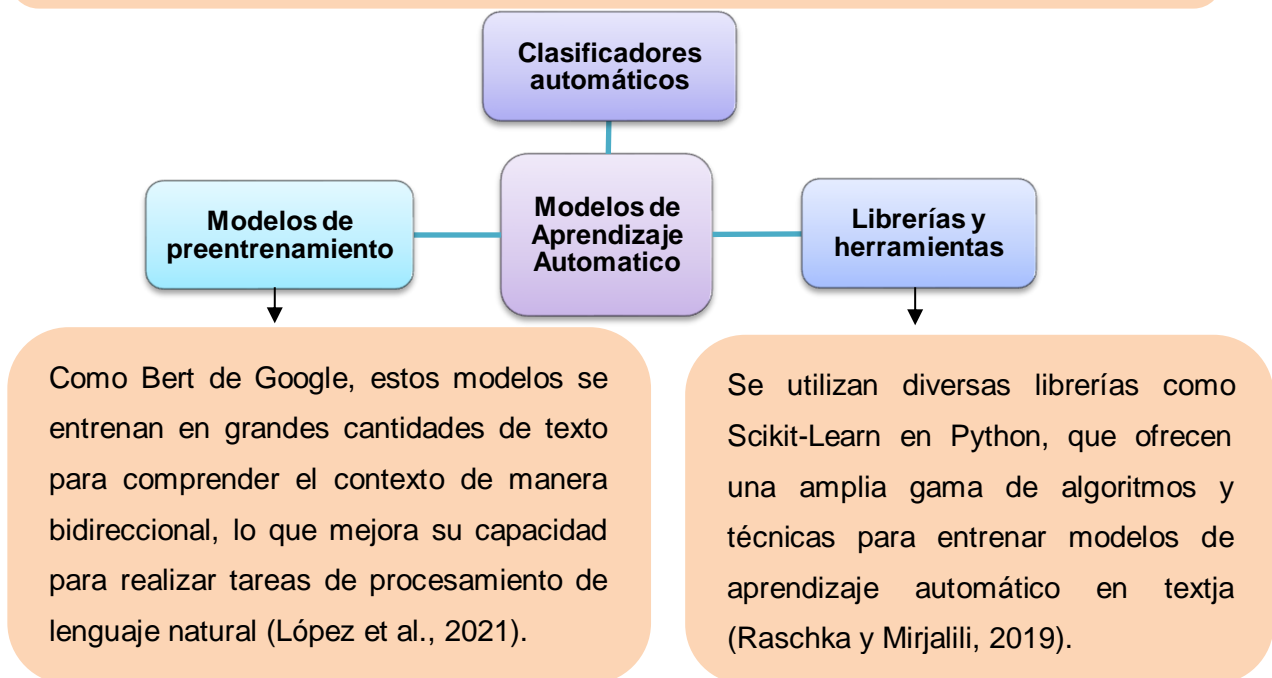
Por otra parte, Guardiola (2020) considera que los modelos de aprendizaje automático para texto son herramientas fundamentales en la actualidad, especialmente en áreas como la clasificación de mensajes en redes sociales y análisis de textos, ya que, estos modelos permiten identificar patrones y características en grandes volúmenes de texto, lo que resulta útil para diversas aplicaciones.

Algunos ejemplos de modelos de aprendizaje automático para texto incluyen:

Figura 6

Modelos de Aprendizaje Automático

Algoritmos que categorizan textos en diferentes clases o categorías, como identificar si un mensaje es racista, habla sobre política o tecnología, o determinar el sentimiento positivo o negativo de un comentario en redes sociales (Pianchiche, 2023).



Nota. Esta es una recopilación de los modelos de aprendizaje automático para texto.

En síntesis, los modelos de aprendizaje automático para texto son cruciales para tareas como clasificación de textos, análisis de sentimientos, identificación de temas y mucho más, utilizando algoritmos avanzados y técnicas de procesamiento de lenguaje natural (Guardiola , 2020).

Los modelos de aprendizaje automático para texto representan una clase especializada de algoritmos diseñados para analizar y comprender datos textuales de manera automatizada, estos modelos son fundamentales en el campo del procesamiento de lenguaje natural, ya que permiten a las computadoras interpretar y extraer información significativa de grandes cantidades de texto de manera eficiente.

2.3 Análisis de sentimientos

El análisis de sentimientos es un proceso donde se determina una opinión sobre un tema en concreto, mediante: valoraciones, actitudes y emociones. Para obtener este objetivo, primero se debe reconocer la expresión y luego definir la orientación del sentimiento expresado, la expresión detectada sobre un tema, servicio, producto o individuo específico puede ser: positiva, negativa o neutra. Puede tener varios enfoques y estar basado en diferentes características como etiquetas, emoticones de uso universal o en recursos léxicos de donde se puedan obtener valoraciones positivas, neutrales o negativas. En Python ya existen algunas librerías con algoritmos desarrollados para el análisis de sentimientos con poder predictivo como por ejemplo la librería MonkeyLearn (Saura et al., 2018)

Cuando nos comunicamos con el resto de personas usamos Lenguaje Natural (LN), gracias a este tipo de comunicación podemos ser altamente expresivos y compartir valiosa información, todas las formas de LN expresadas pueden ser consideradas como datos no estructurados. Actualmente se procesa en gran cantidad estos datos no estructurados en el campo de la informática por medio de técnicas de procesamiento automático. En años recientes ha tomado gran interés el análisis de las redes sociales donde existe una fuente inagotable de opiniones (Cedeño y Vargas, 2020).

El tratamiento automático del lenguaje natural o lenguaje humano se llama PLN, cuyo propósito es analizar la subjetividad en las opiniones o extraer y clasificar los sentimientos de las opiniones expresadas por las personas, en donde el Análisis de sentimientos es el área encargada de la detección automática de los sentimientos expresados en los textos y su clasificación positiva, negativa o neutra. El objetivo del análisis de sentimientos es la extracción de información subjetiva generada por los usuarios en opiniones o blogs (Cedeño y Vargas, 2020).

El análisis de datos es un proceso mediante el cual se examina información no estructurada, como opiniones expresadas en texto, a fin de determinar valoraciones, actitudes y emociones sobre un tema específico, este análisis implica reconocer la expresión del lenguaje natural y clasificar los sentimientos como positivos, negativos o neutros, utilizando herramientas como librerías con algoritmos desarrollados para el análisis de sentimientos, como la librería MonkeyLearn en Python.

2.3.1 Métodos de análisis de sentimientos

Los métodos de análisis de sentimientos según Isasi (2021) son técnicas utilizadas en el procesamiento del lenguaje natural (PLN) para identificar, cuantificar y comprender las opiniones, emociones y actitudes expresadas en textos, estos métodos son ampliamente aplicados en áreas como la minería de opiniones en redes sociales, análisis de comentarios de clientes, evaluación de reseñas de productos y servicios. Algunos de los métodos de análisis de sentimientos más utilizados son:

Tabla 2

Métodos de análisis de sentimientos

Método	Concepto
Detección de Polaridad	Este método se centra en determinar si una opinión o frase es positiva, negativa o neutral. Se basa en la identificación de

	<p>palabras clave, expresiones sentimentales y estructuras gramaticales que denotan emociones. Puede ser implementado mediante el uso de diccionarios de polaridad, aprendizaje automático supervisado (clasificación binaria) o reglas heurísticas (Salaza et al., 2020).</p>
<p>Análisis Basado en Diccionarios</p>	<p>Consiste en el uso de diccionarios léxicos que asignan a cada palabra una puntuación de polaridad (positiva, negativa o neutral) y posiblemente intensidad. Las palabras en el texto se suman para calcular una puntuación global de sentimiento. Ejemplos de diccionarios populares son el Sentiment140, SentiWordNet y VADER (Valence Aware Dictionary and sEntiment Reasoner) (Lovera y Cardinale, 2023).</p>
<p>Aprendizaje Automático Supervisado</p>	<p>Este enfoque implica entrenar modelos de aprendizaje automático, como clasificadores binarios (por ejemplo, Support Vector Machines, Naive Bayes, Redes Neuronales) con datos etiquetados (textos con sentimientos anotados). El modelo aprende a distinguir entre opiniones positivas y negativas en base a características del texto, como palabras clave, sintaxis y contexto (Sánchez et al., 2020).</p>
<p>Análisis de Emociones</p>	<p>A diferencia del análisis de polaridad, el análisis de emociones se centra en identificar emociones específicas, como alegría, tristeza, enojo, miedo, etc. Puede implicar el uso de diccionarios de emociones, técnicas de aprendizaje automático para clasificación multiclase o modelos preentrenados como BERT (Bidirectional Encoder Representations from Transformers) (López y Gonzales, 2021).</p>

Análisis Basado en Aspectos (Aspect-Based Sentiment Analysis)	Este método descompone las opiniones en aspectos específicos y luego evalúa el sentimiento asociado a cada aspecto. Es útil para comprender cómo se siente la gente acerca de diferentes aspectos de un producto, servicio o evento (Quintero y García, 2018).
Redes Neuronales Recurrentes (RNN) y Redes Neuronales Convolucionales (CNN)	Estos modelos de aprendizaje profundo son utilizados para el análisis de sentimientos en textos largos o secuenciales. Las RNN pueden capturar el contexto a lo largo de una secuencia de palabras, mientras que las CNN pueden identificar patrones locales en el texto que indican sentimientos (Sánchez et al., 2020).
Aprendizaje por Transferencia (Transfer Learning)	Se refiere a la técnica de utilizar modelos preentrenados en grandes conjuntos de datos (como BERT, GPT, RoBERTa) y adaptarlos a tareas específicas de análisis de sentimientos. Esto a menudo resulta en un mejor rendimiento, especialmente cuando se dispone de pocos datos etiquetados para entrenar un modelo desde cero (Moreno et al., 2022).

Nota. Esta tabla muestra los métodos de análisis de sentimientos de las AA.

Estos métodos pueden combinarse y ajustarse según las necesidades y el contexto de aplicación del análisis de sentimientos. La elección del método adecuado depende de factores como la disponibilidad de datos etiquetados, la complejidad del texto, la precisión requerida y los recursos computacionales disponibles.

El método VADER (Valence Aware Dictionary for Sentiment Reasoning) es una herramienta ampliamente reconocida para el análisis de sentimientos en textos, su destacado éxito radica en su capacidad para no solo identificar la polaridad positiva o negativa de un texto, sino también para evaluar el grado de positividad o negatividad de los sentimientos expresados, lo cual resulta especialmente útil para investigaciones. Este enfoque se

fundamenta en reglas que clasifican las polaridades de las emociones en las opiniones utilizando un extenso listado de palabras denominado lexicón. Este lexicón comprende un total de 7,517 términos, que incluyen emoticones, abreviaciones, acrónimos e iniciales, todos ellos etiquetados con una valencia que va desde -4 hasta 4. La creación del lexicón de VADER se llevó a cabo mediante la aplicación de metodologías de Machine Learning, la sabiduría de los grupos (Wisdom-of-the-Crowd), y la utilización de otros lexicones reconocidos como LIWC, ANEW y GI (Leyva et al., 2021)

Los métodos de análisis de sentimientos son técnicas utilizadas en el procesamiento de lenguaje tanto natural como de inteligencia artificial para comprender y evaluar la actitud emocional expresada en texto, mediante la aplicación de algoritmos y modelos computacionales para identificar, cuantificar y categorizar las emociones u opiniones de los usuarios expresadas en diferentes tipos de textos, a fin de clasificar el tono del texto como positivo, negativo o neutro, así como también para distinguir emociones específicas como felicidad, tristeza, enojo, entre otras.

2.3.2 Aplicaciones del análisis de sentimientos en redes sociales

Según Pauli (2019), los beneficios derivados del análisis de sentimientos son tanto numerosos como significativos, y se refleja en el interés destacado que muestran empresas, organizaciones e incluso gobiernos a nivel mundial por el desarrollo continuo en este campo de estudio. La capacidad de comprender las opiniones de las personas acerca de productos, políticas y medidas en tiempo real representa una herramienta de gran valor, por lo que, su correcta utilización puede generar ventajas competitivas que antes eran difíciles de alcanzar.

Entre las aplicaciones específicas del análisis de sentimientos, se pueden mencionar las siguientes:

Tabla 3

Aplicaciones del análisis de sentimientos

Tipos de Aplicaciones	Concepto
Valoración de opinión de productos y servicios	Representa una de las aplicaciones más prácticas y directas del análisis de sentimientos, debido a que esta metodología permite a las empresas entender la opinión de los usuarios sobre sus productos sin necesidad de realizar encuestas tradicionales de satisfacción, para así poder determinar si los usuarios tienen una actitud favorable o desfavorable hacia un producto específico, procediendo a ajustar sus estrategias rápidamente para obtener ventajas competitivas (Pauli, 2019).
Posicionamiento de publicidad on-line	Implica que los anunciadores de ciertos productos puedan solicitar que sus anuncios sean mostrados exclusivamente en sitios web que reflejen opiniones positivas, evitando aquellos sitios donde predominen expresiones negativas (Pauli, 2019).
Corrección de opinión	Permite que, en sitios de compras en línea, donde los usuarios dan su opinión junto con una calificación, el sistema de análisis de sentimientos bajo un error del usuario pueda examinar las palabras y corregir automáticamente la calificación incorrecta (Pauli, 2019).
Mejora de los sistemas de recomendación de productos	Esta metodología permite que, basándose en las opiniones de los usuarios, una tienda en línea pueda decidir destacar los productos que tienen opiniones favorables o evitar

	aquellos que tienen una opinión general negativa (Pauli, 2019).
Reputación política	El análisis de sentimientos por medio de un algoritmo permite reconocer la opinión de la gente sobre un determinado grupo político o postulante gubernamental (Pauli, 2019).
Análisis del mercado financiero	Analizando la información recopilada de páginas web, foros y redes sociales, es factible anticipar el desempeño en el mercado financiero considerando el valor añadido de la polaridad de todas las opiniones recopiladas (Pauli, 2019).

Nota. Esta tabla muestra las principales aplicaciones para el análisis de sentimientos.

El análisis de sentimientos en redes sociales tiene una relevancia creciente en la actualidad, dado el enorme volumen de datos generados por usuarios en plataformas digitales, permitiendo comprender las opiniones, emociones y actitudes de los usuarios hacia diversos temas, productos o eventos, ya que al aplicar el análisis de sentimientos en redes sociales, se puede identificar tendencias, evaluar la percepción pública sobre una marca o tema específico, y tomar decisiones estratégicas basadas en la retroalimentación directa de los usuarios.

2.4 Herramientas y tecnologías

Las herramientas y tecnologías se refieren a los recursos, instrumentos y sistemas utilizados para llevar a cabo diferentes tareas en diversos campos, como la informática, la ingeniería, la medicina, entre otros. Estas herramientas y tecnologías pueden ser tanto físicas como virtuales, y su función radica tanto en facilitar como en mejorar la ejecución de actividades específicas, algunas de estas herramientas más utilizadas son Python y Google Colab (Llerena et al., 2024).

2.4.1 Python y google colab

Python es un lenguaje de programación de alto nivel, interpretado y de propósito general, que se destaca por su sintaxis clara y legible, lo que lo hace ideal para desarrolladores de diferentes niveles de experiencia, además este lenguaje es conocido por su versatilidad y su amplio uso en diversas áreas, como desarrollo web, análisis de datos, inteligencia artificial, aprendizaje automático y ciencia de datos. Además, cuenta con una amplia comunidad de usuarios y una gran cantidad de bibliotecas o herramientas que facilitan el desarrollo tanto de aplicaciones como de proyectos de software (Vidal et al., 2021).

Google Colaboratory (también conocido como Google Colab) es un servicio en la nube basado en Jupyter Notebooks para difundir la educación y la investigación del aprendizaje automático. Proporciona un tiempo de ejecución totalmente configurado para aprendizaje profundo y acceso gratuito a una GPU o TPU robusta permitiendo compartir el código de Python de forma colaborativa. Google Colaboratory está alojada en la plataforma Google Cloud, los cuadernos de Colaboratory funcionan como un objeto de Google Docs, se puede compartir y varios usuarios pueden colaborar en el mismo cuaderno (Carneiro et al., 2018).

Python es un lenguaje de programación versátil que permite crear programas complejos con una sintaxis clara y legible, que permite traducir ideas en código de manera eficiente. Por otro lado, Google Colab es un entorno de desarrollo que permite ejecutar y compartir el código de Python. Además, está basado en Jupyter Notebooks y proporciona acceso gratuito a recursos computacionales, como GPU y TPU, lo que lo convierte en una opción popular para proyectos que requieren un alto poder de procesamiento.

2.4.2 React JS

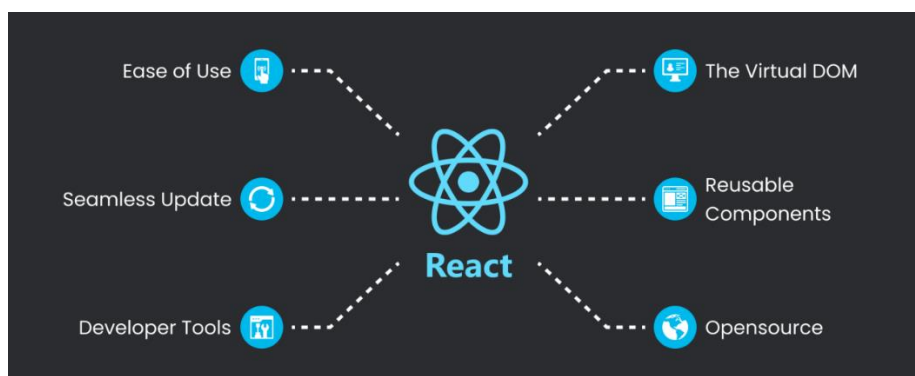
React JS es una biblioteca de JavaScript de código abierto desarrollada por Facebook que se utiliza principalmente para crear interfaces de usuario interactivas y de alta calidad, esta utiliza un enfoque basado en componentes, lo que significa que las diferentes partes de una interfaz de usuario se representan como componentes reutilizables y modulares, lo que

facilita la construcción de aplicaciones complejas al dividir las en pequeños bloques de código que se pueden desarrollar, probar o incluso mantener de manera independiente. Una de las características clave de React JS es su capacidad para actualizar eficientemente la interfaz de usuario cuando los datos cambian, empleando un concepto llamado "Virtual DOM" (DOM virtual) que permite minimizar los cambios en el DOM real, lo que resulta en un rendimiento más rápido y una experiencia de usuario más fluida (Aggarwal, 2018).

El React JS al ser una biblioteca de Java fomenta el uso de JSX (JavaScript XML), una extensión de JavaScript que permite escribir código HTML dentro de JavaScript, lo que facilita la creación de componentes visualmente atractivos y dinámicos, así como también permite la incorporación de lógica de programación directamente en la estructura de la interfaz de usuario. Otra primacía de React JS es su amplio ecosistema de herramientas y bibliotecas complementarias, como React Router para el enrutamiento, Redux para la gestión del estado, y Material-UI o Bootstrap para estilos y componentes predefinidos, haciendo que sea una opción popular para el desarrollo web, especialmente en aplicaciones de una sola página (SPA) y aplicaciones web progresivas (PWA) que requieren interactividad y rendimiento optimizado (Syah y Anggraini, 2022).

Figura 7

Imagen de React JS



Nota. Google Scholar. (s. f.). React JS components (Fotografía). Recuperado el 18 de abril de 2024, de <https://www.linkedin.com/pulse/reactjs-ecosystem-exploring-key->

El React JS es una biblioteca de JavaScript diseñada para crear interfaces de usuario de intercambio colaborativo que utiliza componentes reutilizables y un enfoque basado en el Virtual DOM para actualizar la interfaz de manera ágil y fluida. Su uso de JSX simplifica la creación de elementos visuales y la incorporación de lógica de programación, lo que lo hace ideal para el desarrollo web de aplicaciones dinámicas y de alto rendimiento.

2.4.3 Librerías y frameworks para aprendizaje automático

Según Nguyen (2019) existen varias librerías y frameworks ampliamente utilizados en el campo del aprendizaje automático que ofrecen herramientas y funcionalidades avanzadas para el desarrollo de modelos predictivos y análisis de datos. Algunas de las más destacadas son:

Tabla 4

Ejemplos de librerías y frameworks

Ejemplos	Concepto
Scikit-Learn	Es una librería ampliamente utilizada para el aprendizaje automático tradicional, como la clasificación, regresión, clustering y preprocesamiento de datos, además proporciona una amplia variedad de algoritmos de aprendizaje automático listos para usar, junto con herramientas para evaluar y optimizar el rendimiento de los modelos (Nguyen, y otros, 2019)
TensorFlow	Desarrollada por Google, es una de las librerías más populares para el aprendizaje automático y el desarrollo de modelos de inteligencia artificial. Ofrece una amplia gama de herramientas para construir redes neuronales, incluyendo redes neuronales convolucionales (CNN) y recurrentes (RNN), así como también soporte para el aprendizaje

	profundo (deep learning) y técnicas de aprendizaje por transferencia (Transfer Learning) (Nguyen, y otros, 2019).
TensorFlow.js	Dirigida al desarrollo de aplicaciones de aprendizaje automático en el navegador y en entornos de Node.js, esta permite implementar modelos de TensorFlow en JavaScript, lo que facilita la creación de aplicaciones interactivas basadas en la web que utilizan inteligencia artificial (Nguyen, y otros, 2019).
Keras	Es una librería de alto nivel que se integra con TensorFlow y proporciona una interfaz fácil de usar tanto para construir como para entrenar modelos de aprendizaje profundo, es conocida por su simplicidad al igual que su flexibilidad, lo que la hace ideal para experimentar rápidamente con diferentes arquitecturas de redes neuronales (Nguyen, y otros, 2019).
Microsoft Cognitive Toolkit (CNTK)	Desarrollada por Microsoft, CNTK es una librería de código abierto que ofrece herramientas para el entrenamiento de modelos de aprendizaje profundo, especialmente diseñada para aplicaciones de procesamiento del lenguaje natural (NLP) y visión por computadora, se destaca por su eficiencia en el entrenamiento de modelos complejos en entornos distribuidos (Nguyen, y otros, 2019). Caffe: Es un framework desarrollado por Berkeley AI Research (BAIR) que se centra en la velocidad, además esta optimizado para tareas de visión por computadora y redes neuronales convolucionales (CNN), siendo ampliamente utilizado en aplicaciones de reconocimiento de imágenes y videos (Nguyen, y otros, 2019).
Caffe2	Es la versión sucesora de Caffe, que está diseñado para ofrecer una mayor flexibilidad al mismo tiempo que eficiencia en el desarrollo y

	despliegue de modelos de aprendizaje profundo en diversos dispositivos y plataformas, incluyendo dispositivos móviles y sistemas embebidos (Nguyen, y otros, 2019).
Torch	Es un framework de aprendizaje automático - aprendizaje profundo que ofrece una interfaz sencilla para la creación y entrenamiento de modelos, así como también herramientas para la experimentación y visualización de resultados (Nguyen, y otros, 2019).
PyTorch	Desarrollada por Facebook, es una librería popular para el aprendizaje automático que se destaca por su capacidad de construir modelos de manera dinámica, lo que facilita la experimentación y la depuración de código, además ofrece soporte para el cálculo en GPUs, lo que acelera significativamente el entrenamiento de modelos complejos (Nguyen, y otros, 2019).
MXNet	Desarrollado por Apache Software Foundation, es un framework escalable, eficiente para el aprendizaje automático distribuido, que ofrece soporte para múltiples lenguajes de programación, incluyendo Python, R, Scala y Julia, que se destaca por su capacidad para trabajar con conjuntos de datos de gran tamaño (Nguyen, y otros, 2019).
Chainer	Es un framework de aprendizaje profundo que se caracteriza por su capacidad para definir modelos de manera dinámica, permitiendo la construcción de modelos más "imperativos", facilitando la experimentación y el desarrollo de arquitecturas personalizadas (Nguyen, y otros, 2019).
Theano	Fue una librería popular en el cálculo numérico y la construcción de modelos de aprendizaje automático, que se destacó por su capacidad para optimizar el rendimiento computacional mediante la compilación de

	expresiones matemáticas en código eficiente, especialmente para el entrenamiento de redes neuronales (Nguyen, y otros, 2019).
Performance-wise preliminary y deep learning wrapper libraries	Se centran en optimizar tanto el rendimiento como la eficiencia de los modelos de aprendizaje profundo en hardware específico, como GPUs o CPUs, que ofrecen herramientas para acelerar el entrenamiento y la inferencia de modelos, aprovechando al máximo las capacidades de cómputo de las plataformas hardware (Nguyen, y otros, 2019).
Freeling	Es una librería de código abierto para el análisis de texto para distintos idiomas, su instalación no es sencilla y es aún más complicada en el sistema operativo de Windows, pero una vez resuelto este inconveniente se puede utilizar todas las funciones implementadas (Padro y Stanilovsky, 2022).

Nota. Esta tabla muestra ejemplos de algunas librerías y frameworks.

Las librerías y frameworks son herramientas que facilitan el desarrollo, entrenamiento y despliegue de modelos de inteligencia artificial, permitiendo a los desarrolladores implementar algoritmos complejos, como redes neuronales y modelos de aprendizaje profundo, de manera tanto eficiente como escalable. Estas herramientas ofrecen funcionalidades avanzadas, como la optimización del rendimiento, la interoperabilidad con diferentes lenguajes de programación y la capacidad de trabajar en entornos distribuidos.

2.5 Trabajos relacionados

2.5.1 Estudios sobre clasificación de textos en redes sociales

Según Gómez et al. (2018) en su artículo “Influencia de redes sociales en el análisis de sentimiento aplicado a la situación política en Ecuador” se observó que hay una gran relación entre las plataformas digitales como Twitter y la situación política del país basado en un análisis de sentimientos de los usuarios en donde expresan sus preferencias y opiniones,

por lo que, aplicando el método de polarización en datos reales obtenidos de la plataforma antes mencionada, se pudo identificar tendencias, evaluar la percepción pública y detectar posibles patrones de comportamiento en la opinión política, dando como resultado que las elecciones Ecuador 2017 no coincidieron con el análisis de datos realizado en Twitter, en consecuencia se estableció que la relación entre tendencias en redes de microblogging y los resultados electorales incluyen más variables de las que percibimos, por lo que los algoritmos basado en el aprendizaje automatizado deberían abarcar una mayor área de relacionamiento.

Por otro lado, Burdisso et al. (2019) en su estudio denominado “Un marco de clasificación de textos para la detección temprana de la depresión simple y eficaz en las redes sociales” evidenció que la clasificación de textos en redes sociales juega un papel crucial en la detección temprana de la depresión, puesto que permite identificar patrones lingüísticos y emocionales, involucrando el uso de algoritmos de aprendizaje automático, como las redes neuronales convolucionales (CNN) o los modelos de procesamiento de lenguaje natural (NLP), que pueden analizar grandes cantidades de datos en tiempo real, dando como resultado la identificación de características específicas, como el tono emocional, la frecuencia de palabras relacionadas con la depresión y la estructura gramatical, a fin de distinguir mensajes que podrían indicar depresión de manera temprana.

Según Wu et al. (2018) en su análisis “Rastreado las huellas de las noticias falsas: caracterizando los mensajes de las redes sociales según cómo se propagan” se identificó que la clasificación de textos en redes sociales, especialmente en el contexto de rastrear las huellas de las noticias falsas y caracterizar los mensajes según cómo se propagan, es un desafío clave en la era digital. Por ello, al emplear técnicas avanzadas de procesamiento de lenguaje natural y aprendizaje automático como el TraceMiner y el LSTM-RNN para analizar el contenido de los mensajes y determinar tanto su veracidad como su nivel de influencia, permiten identificar patrones lingüísticos, analizar la difusión de información a través de redes sociales y evaluar la credibilidad de las fuentes con un alto grado de precisión. Además, al

combinar algoritmos de clasificación de texto con modelos de análisis de sentimiento, permiten detectar mensajes engañosos o manipulados y comprender cómo afectan la percepción pública y el debate en línea, lo que es fundamental para mitigar los efectos negativos de la desinformación y promover un entorno digital más confiable y transparente.

Entorno al sector de marketing, Moreno et al. (2020) en su investigación “Técnicas para la Clasificación de Sentimientos en Redes Sociales como Apoyo en el Marketing Digital” evidenció que la clasificación de sentimientos en redes sociales ha emergido como una herramienta poderosa mediante técnicas de minería de datos como el algoritmo SVM, que se centra en analizar grandes volúmenes de datos para identificar patrones de comportamiento de sus usuarios, a fin de medir la aceptación de sus productos, propuestas y campañas de marketing, lo que les permite tomar decisiones informadas. Esto les permite a las empresas tener una percepción clara de las preferencias del cliente, identificar tendencias, gestionar proactivamente la marca y de ser el caso segmentar a la audiencia, mejorando la experiencia del cliente y optimizando el impacto de sus acciones de marketing en el mercado.

En cuanto a Recalde (2021), en su estudio “El ciberacoso por redes sociales en el Ecuador” menciona que el acoso cibernético es un problema creciente, donde diversas plataformas son utilizadas por personas de todas las edades y contextos sociales, que pueden causarse daños emocionales y afectaciones en la vida tanto de los vulnerados como de sus entornos. En un análisis actual se evidencio un incremento alarmante en denuncias por violación a la intimidad (18%) y pornografía adolescente (32%), por lo que implementar estrategias de clasificación de textos se ha vuelto esencial. Basado en esto, al emplear algoritmos como el Machine Learning (SVM, regresión logística), se puede detectar rápidamente contenido dañino, priorizar casos urgentes, personalizar respuestas y analizar tendencias para desarrollar estrategias efectivas de prevención y educación, lo que da como resultado la identificación y categorización de mensajes, facilitando medidas como la eliminación de publicaciones ofensivas y el bloqueo de usuarios infractores.

La importancia de los estudios sobre clasificación de textos en redes sociales radica en su capacidad tanto para analizar como para comprender el vasto volumen de información textual generada en plataformas digitales, estos estudios son primordiales para la detección de tendencias, el análisis de opiniones y la identificación de patrones de comportamiento (bueno o malo), lo que permite a empresas e investigadores tomar decisiones informadas, desarrollar estrategias efectivas o entender mejor las dinámicas sociales en línea de acuerdo a cada contexto.

2.5.2 Avances recientes en aprendizaje automático aplicado a redes sociales

Las redes sociales digitales han popularizado la comunicación interactiva, permitiendo compartir historias y experiencias en diversos formatos, no obstante, junto con esta apertura, han surgido comportamientos anómalos, como el trolling, que puede manifestarse de diversas formas, desde irrespeto y sarcasmo hasta cyberbullying o desinformación. Para abordar este problema, se han desarrollado metodologías basadas en minería de datos y aprendizaje automático, identificando algoritmos como Random Forest, Regression Lineal, Support Vector Machine, Naïve Bayes, Kmeans, SMO, Gradient Boost y CNN como herramientas utilizadas para detectar el trolling. A pesar de los avances, sigue siendo un desafío obtener una certeza alta en la clasificación del contenido troll y en la detección de perfiles de trolls, debido a la privacidad y a las variaciones semánticas del lenguaje. Sin embargo, el progreso en la identificación del trolling es evidente, mostrando el potencial del aprendizaje automático para abordar problemas complejos en el entorno digital de las redes sociales (Carrión y Segovia, 2022).

Los avances en el aprendizaje automático aplicado a redes sociales, especialmente en el análisis emocional, son cada vez más significativos, por ejemplo, en la actualidad, la comunicación por medios telemáticos y redes sociales es muy común, lo que ha generado cambios en la forma de expresión de las personas, en este contexto, los modelos de deep learning, como los transformers y específicamente BERT, han demostrado ser herramientas

poderosas para el procesamiento natural del lenguaje, como la librería transformers, desarrollada por la comunidad de inteligencia artificial Hugging Face, que ha permitido un análisis más profundo del contexto del mensaje, especialmente en tareas como el análisis de polaridad y emociones. Además, estudios recientes han logrado precisiones superiores al 95% en el análisis de polaridad utilizando modelos de transformers, lo que evidencia su efectividad y precisión en la comprensión del tono emocional en mensajes dentro del entorno de las redes sociales, estos avances representan un hito significativo en la capacidad de las tecnologías de aprendizaje automático cuyo fin es entender y procesar la complejidad de la comunicación digital en la era actual (Estévez, 2023).

En los últimos años, se han producido avances significativos en el campo del aprendizaje automático aplicado a redes sociales, centrándose en el análisis de sentimiento, la detección de temas emergentes, la identificación de usuarios influyentes y la personalización de contenido. En este sentido, los modelos de aprendizaje profundo han permitido mejorar la precisión en la clasificación de textos en redes sociales, especialmente en la detección de emociones y opiniones. Además, han facilitado la adaptación de modelos pre entrenados a contextos específicos de redes sociales, optimizando así el rendimiento de los sistemas de aprendizaje automático. Estos avances han contribuido a la comprensión de la dinámica de las interacciones en redes sociales al mismo tiempo que han abierto nuevas oportunidades para la aplicación de técnicas de aprendizaje automático en la mejora de la experiencia del usuario y la toma de decisiones en entornos digitales (Díaz y Suárez, 2019).

Los avances recientes en aprendizaje automático aplicado a redes sociales se centran en técnicas avanzadas de procesamiento de lenguaje natural (PLN) para mejorar la comprensión de textos, como el análisis de sentimientos y la detección de temas emergentes. Además, se están desarrollando modelos de aprendizaje profundo que permiten una mejor personalización de contenidos y recomendaciones, así como la detección de contenido inapropiado o fraudulento de manera más efectiva. Estos avances están mejorando

significativamente la capacidad de las plataformas de redes sociales para ofrecer experiencias más relevantes y seguras para los usuarios (Salgado , 2024).

Los avances en aprendizaje automático aplicado a redes sociales se enfocan en mejorar la comprensión del texto mediante técnicas avanzadas de procesamiento de lenguaje natural, como el análisis de sentimientos y la detección de temas emergentes, con el fin no solo de ayudar a entender mejor las interacciones en redes sociales, sino también de crear oportunidades para mejorar la experiencia del usuario y la toma de decisiones en entornos digitales de manera más relevante y segura, por ello, los modelos han aumentado la precisión en la clasificación de textos, permitiendo adaptar modelos a contextos específicos de redes.

Capítulo tres

Metodología

3.1 Diseño de la investigación

Esta investigación sigue una metodología aplicada y descriptiva para analizar el sentimiento en las publicaciones en las redes sociales mediante la aplicación técnicas de aprendizaje automático. El enfoque se basa en el proceso de Descubrimiento de Conocimiento en Bases de Datos (KDD, del inglés Knowledge Discovery in Databases), que implica una serie de pasos para extraer conocimiento valioso de los datos. El carácter aplicado de este estudio pretende aportar soluciones prácticas para el análisis de sentimientos, mientras que el aspecto descriptivo se centra en detallar el rendimiento y las características de varios modelos de aprendizaje automático.

Figura 8

Imagen del Proceso KDD



El proceso de KDD comienza con la selección de los datos pertinentes, seguida del pre procesamiento, la transformación, la extracción de datos y la interpretación. Cada paso es fundamental para garantizar que los datos se preparen con precisión para el análisis y que

los resultados son fiables y significativos. Este enfoque estructurado permite un examen exhaustivo de las técnicas de análisis de los sentimientos, lo que proporciona información sobre la eficacia de los distintos modelos y metodologías en un contexto práctico.

El diseño de la investigación también incorpora un análisis comparativo de múltiples algoritmos de aprendizaje automático, en concreto Naive Bayes, Support Vector Machine (SVM) y Random Forest, para determinar el modelo más eficaz para la clasificación de sentimientos. Al seguir el marco KDD, este estudio garantiza una metodología sistemática y replicable que puede aplicarse a otros conjuntos de datos y dominios, lo que mejora su validez y aplicabilidad general.

3.2 Recopilación de datos

Los datos para esta investigación se recopilaron a partir de conjuntos de datos públicos disponibles en Kaggle, centrándose específicamente en las publicaciones de redes sociales de Twitter; en algún caso, de manera adicional, se encuentra la red social Reddit. El conjunto de datos incluye varias características, como texto, información del usuario y metadatos, que son cruciales para realizar análisis de sentimientos. La selección de este conjunto de datos se basó en su relevancia y exhaustividad, y proporciona una base sólida para entrenar y evaluar modelos de aprendizaje automático.

En primera instancia se descarga un dataset propio de Kaggle llamado Cosmos 98 el cual contiene publicaciones de Twitter y de Reddit. A continuación, una muestra de los datos que se obtiene de estos datasets:

Figura 9

Imagen de ejemplo de datasets

index	clean_text	category
0	when modi promised "minimum government maximum governance" expected him begin the difficult job reforming the state why does take years get justice state should and not business and should exit psus and temples	-1.0
1	talk all the nonsense and continue all the drama will vote for modi	0.0
2	what did just say vote for modi welcome bjp told you rahul the main campaigner for modi think modi should just relax	1.0
3	asking his supporters prefix chowkidar their names modi did great service now there confusion what read what not now crustal clear what will crass filthy nonsensical see how most abuses are coming from chowkidars	1.0
4	answer who among these the most powerful world leader today trump putin modi may	1.0

Show | 25 | per page

El investigar varios datasets públicos sirvió para familiarizarse con la forma de obtención de datos. Luego se procede a descargar otro dataset llamado Sentiment140 también descargado de Kaggle, el cual contiene 1,6 millones tweets.

El proceso de recopilación de datos también implicó garantizar la calidad e integridad del conjunto de datos. Para ello, se comprobó si faltaban valores, si había duplicados o incoherencias que pudieran afectar al análisis. Al utilizar un conjunto de datos de acceso público, la investigación respeta las normas éticas y garantiza que los datos sean representativos de las interacciones típicas en las redes sociales.

Además, el conjunto de datos se dividió en conjuntos de entrenamiento y de prueba para facilitar el desarrollo y la evaluación del modelo. Esta división es esencial para validar el rendimiento de los modelos y garantizar que los resultados no estén sesgados o sobre ajustados a los datos de entrenamiento. El uso de un conjunto de datos bien seleccionados de una fuente de confianza, como Kaggle, garantiza la fiabilidad de los datos y la validez de los resultados de la investigación.

3.3 Preprocesamiento de datos

El preprocesamiento de datos es un paso crucial en la metodología KDD, cuyo objetivo es mejorar la calidad y fiabilidad de los datos. En este proyecto, el preprocesamiento implicó varias tareas clave: limpieza, normalización y transformación de los datos. Inicialmente, eliminamos los mensajes duplicados, tratamos los valores que faltaban y filtramos cualquier contenido irrelevante para garantizar la integridad del conjunto de datos.

La normalización consistió en convertir todo el texto a minúsculas, eliminar los signos de puntuación, los números y las palabras vacías, que son palabras comunes que no aportan un significado significativo al texto. Este paso es esencial para reducir el ruido en los datos y garantizar que el texto tenga un formato coherente para el análisis. A continuación, se procedió a la tokenización para dividir el texto en palabras individuales o tokens, que son las unidades básicas de análisis en el procesamiento del lenguaje natural (PLN).

La fase de transformación incluyó la conversión de los datos de texto en representaciones numéricas mediante técnicas como la Frecuencia de Términos-Frecuencia Inversa de Documentos (TF-IDF). Este método ayuda a captar la importancia de las palabras en el contexto de todo el conjunto de datos. Al transformar el texto en vectores, lo hacemos apto para su introducción en algoritmos de aprendizaje automático. Estos pasos de preprocesamiento son vitales para garantizar que los datos estén limpios, sean coherentes y estén listos para el entrenamiento del modelo.

3.4 Modelos de clasificación

En la fase de minería de datos del proceso KDD, nos centramos en construir y entrenar varios modelos de aprendizaje automático para clasificar el sentimiento de las publicaciones de redes sociales seleccionamos tres algoritmos de clasificación populares: Naive Bayes, Support Vector Machines (SVM) y Random Forest. Cada uno de estos modelos tiene puntos fuertes únicos y se utiliza ampliamente en tareas de análisis de sentimientos.

Naive Bayes es un clasificador probabilístico que aplica el teorema de Bayes, lo que lo hace sencillo y eficaz para tareas de clasificación de texto. SVM, por otro lado, es un potente clasificador lineal que funciona bien con datos de alta dimensión y es eficaz para capturar las complejas relaciones entre características. Random Forest es un método de aprendizaje conjunto que combina varios árboles de decisión para mejorar la precisión de la clasificación y reducir el sobreajuste.

Cada modelo se entrenó utilizando el conjunto de datos preprocesados y se evaluó en un conjunto de pruebas independiente. El rendimiento de cada modelo se evaluó utilizando métricas como la precisión, la recuperación, la puntuación F1 y la exactitud. Esta evaluación exhaustiva ayuda a comprender los puntos fuertes y débiles de cada algoritmo en el contexto del análisis de sentimientos. A continuación, se seleccionó el modelo con mejores resultados para su posterior desarrollo e implantación.

Capítulo cuatro

Desarrollo del proyecto

4.1 Desarrollo del modelo de clasificación

El desarrollo del modelo de clasificación consistió en entrenar los algoritmos de aprendizaje automático seleccionados en el conjunto de datos preprocesados de publicaciones de Twitter. Utilizando Python y bibliotecas como scikit-learn, implementamos los clasificadores Naive Bayes, SVM y Random Forest. El conjunto de datos se dividió en conjuntos de entrenamiento y de prueba en una proporción de 80-20 para garantizar que los modelos se evaluaran con datos no vistos.

Durante la fase de entrenamiento, cada modelo se optimizó mediante validación cruzada para ajustar los parámetros y mejorar el rendimiento. Por ejemplo, el modelo SVM se afinó con diferentes funciones de núcleo, mientras que el número de árboles del modelo Random Forest se ajustó para equilibrar la precisión y la eficiencia computacional. Este proceso iterativo de entrenamiento y validación garantizó que los modelos fueran sólidos y capaces de manejar datos del mundo real.

Tras el entrenamiento, los modelos se guardaron utilizando la biblioteca pickle de Python, que permite almacenarlos y recuperarlos fácilmente. Este paso permite reutilizar los modelos sin necesidad de volver a entrenarlos, lo que ahorra tiempo y recursos informáticos. A continuación, los modelos guardados se probaron en el conjunto de datos de prueba para confirmar su rendimiento y fiabilidad.

4.2 Evaluación del modelo

La evaluación de modelos es un paso fundamental para garantizar que los modelos de análisis de sentimientos funcionan bien con datos no vistos. Las métricas de evaluación utilizadas fueron la exactitud, la precisión, la recuperación, la puntuación F1 y el soporte. Estas métricas proporcionan una visión completa del rendimiento de los modelos desde distintos ángulos. La exactitud mide la corrección general, mientras que la precisión y la

sensibilidad se centran en la capacidad de los modelos para identificar correctamente los sentimientos positivos y negativos, respectivamente. La puntuación F1 proporciona un equilibrio entre la precisión y la sensibilidad, y el soporte indica el número de casos verdaderos para cada clase.

La precisión es el porcentaje de clasificaciones correctas de nuestro clasificador, de todo lo que nuestro clasificador clasifica como positivo correcta o incorrectamente.

$$\text{Precisión} = \text{verdaderos positivos} / (\text{Verdaderos positivos} + \text{falsos positivos})$$

El recall o sensibilidad de nuestro modelo es el porcentaje de positivos detectado en el dataset por nuestro clasificador.

$$\text{Recall} = \text{verdaderos positivos} / (\text{Verdaderos positivos} + \text{falsos negativos})$$

Los resultados de la evaluación muestran que cada modelo tiene sus puntos fuertes. Naive Bayes funcionó bien con un conjunto de datos equilibrado, mostrando una precisión y una recuperación elevadas.

Con el 5% del dataset los resultados fueron los siguientes:

Figura 10

Resultados con el 5% dataset

```

Multinomial Naive Bayes
Accuracy: 0.7439375
      precision  recall  f1-score  support
0         0.74    0.75    0.74    8005
1         0.74    0.74    0.74    7995

accuracy                0.74    16000
macro avg              0.74    0.74    0.74    16000
weighted avg          0.74    0.74    0.74    16000

```

Por otro lado, con el 20% del dataset los resultados fueron los siguientes:

Figura 11*Resultados con el 20% dataset*

```

Multinomial Naive Bayes
Accuracy: 0.751046875
      precision    recall  f1-score   support

     0       0.76     0.74     0.75     31912
     1       0.75     0.76     0.75     32088

 accuracy                0.75     64000
 macro avg       0.75     0.75     0.75     64000
 weighted avg    0.75     0.75     0.75     64000

```

SVM sobresalió en el manejo de datos de alta dimensión, proporcionando alta precisión y puntuaciones F1.

Con el 5% del dataset los resultados fueron los siguientes:

Figura 12*Resultados con el 5% dataset – SVM*

```

Support Vector Machine
Accuracy: 0.7543125
      precision    recall  f1-score   support

     0       0.76     0.74     0.75     8005
     1       0.75     0.77     0.76     7995

 accuracy                0.75     16000
 macro avg       0.75     0.75     0.75     16000
 weighted avg    0.75     0.75     0.75     16000

```

Mientras que con el 20% del dataset los resultados fueron los siguientes:

Figura 13

Resultados con el 20% dataset - SVM

```
Support Vector Machine
Accuracy: 0.7843125
      precision    recall  f1-score   support

     0       0.79      0.77      0.78     32020
     1       0.78      0.80      0.79     31980

 accuracy                   0.78     64000
 macro avg                   0.78     64000
 weighted avg                 0.78    640000
```

Random Forest demostró robustez y versatilidad, logrando un alto rendimiento en todas las métricas.

Con el 5% del dataset los resultados fueron los siguientes:

Figura 14

Resultados con el 5% dataset – Random Forest

```
Random Forest
Accuracy: 0.7411875
      precision    recall  f1-score   support

     0       0.74      0.73      0.74      8005
     1       0.74      0.75      0.74      7995

 accuracy                   0.74     16000
 macro avg                   0.74     16000
 weighted avg                 0.74     16000
```

Con el 20% del dataset los resultados fueron los siguientes:

Figura 15

Resultados con el 20% dataset – Random Forest

```

Random Forest
Accuracy: 0.75075
      precision    recall  f1-score   support

     0       0.75     0.74     0.75     31912
     1       0.75     0.76     0.75     32088

 accuracy          0.75     64000
 macro avg         0.75     0.75     0.75     64000
 weighted avg      0.75     0.75     0.75     64000

```

4.3 Visualización de resultados

Para visualizar los resultados, generamos varios diagramas y gráficos utilizando bibliotecas como Matplotlib y Seaborn. Se utilizaron matrices de confusión para ilustrar los verdaderos positivos, falsos positivos, verdaderos negativos y falsos negativos de cada modelo. Estas matrices ayudaron a identificar cualquier error de clasificación y a comprender la distribución de los errores. Además, se trazaron curvas de precisión-sensibilidad y curvas ROC (Receiver Operating Characteristic) para proporcionar información sobre el rendimiento de los modelos en diferentes niveles de umbral. Una curva ROC es una representación gráfica del rendimiento de un modelo de clasificación binaria para todos los umbrales de clasificación.

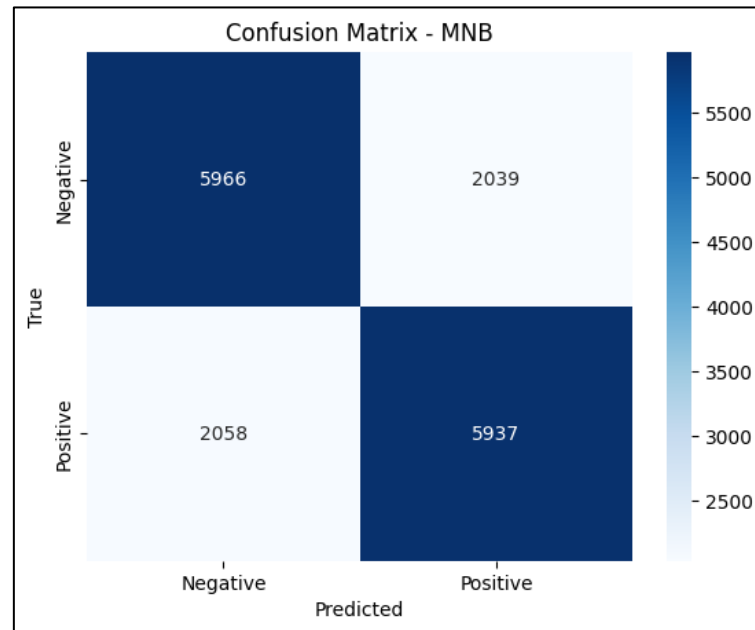
Estas visualizaciones facilitaron la comparación de los modelos y la comprensión de sus puntos fuertes y débiles. Las representaciones gráficas de las métricas de evaluación proporcionaron una forma clara e intuitiva de comunicar los resultados del análisis de sentimientos.

Con el 5% del dataset que son 16000 datos, en el modelo Naive Bayes podemos observar en la siguiente matriz de confusión que existieron 5966 verdaderos negativos y 5937

verdaderos positivos, por otro lado, en los cuadrantes de color blanco vemos que se obtuvo 2039 falsos negativos y 2058 falsos positivos.

Figura 16

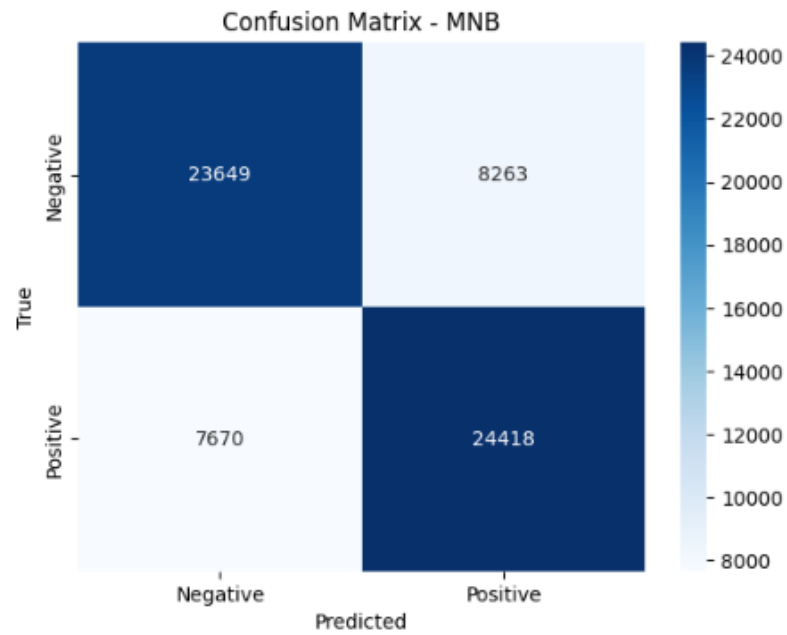
Matriz de confusión – MNB con el 5% del dataset



Con el 20% del dataset que son 64000 datos, en el modelo Naive Bayes podemos observar en la siguiente matriz de confusión que existieron 23649 verdaderos negativos y 24418 verdaderos positivos, por otro lado, en los cuadrantes de color blanco vemos que se obtuvo 8263 falsos negativos y 7670 falsos positivos.

Figura 17

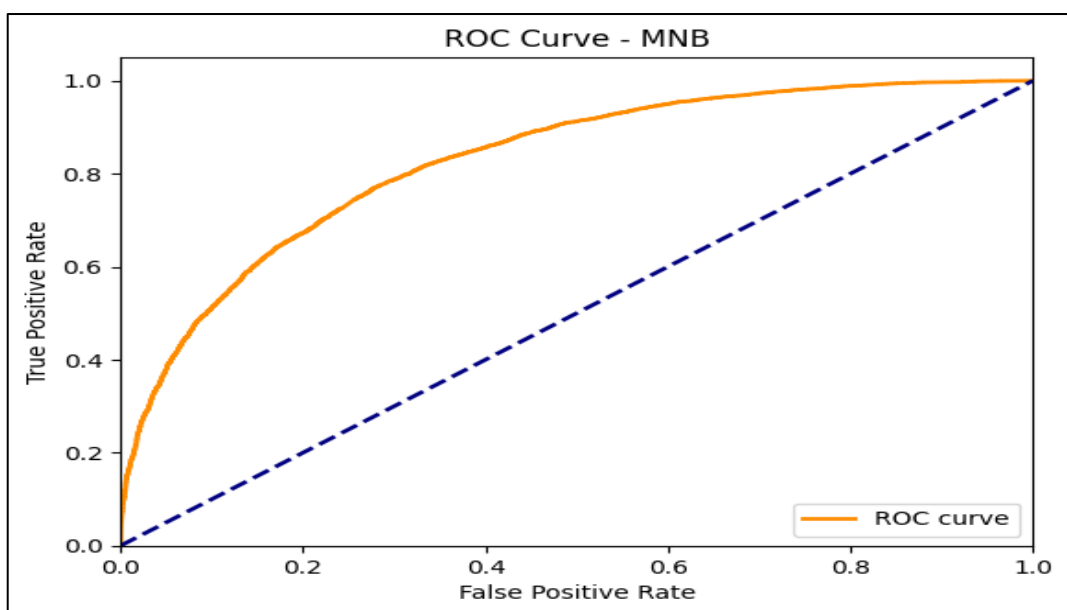
Matriz de confusión – MNB con el 20% del dataset



Además, vemos en el modelo Naive Bayes por medio de la curva ROC la Tasa de Verdaderos Positivos representados en el eje y y la Tasa de Falsos Positivos en el eje x.

Figura 18

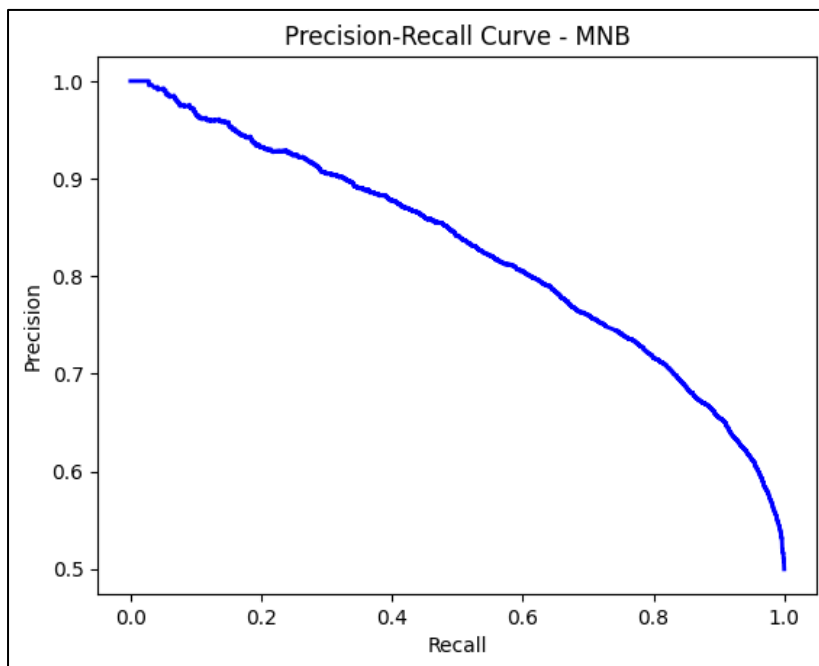
Curva ROC - MNB



Por último, en el modelo Naive Bayes en la curva de precisión-sensibilidad podemos ver el comportamiento de estos dos parámetros, en donde siempre buscaremos tener una alta precisión y una alta sensibilidad.

Figura 19

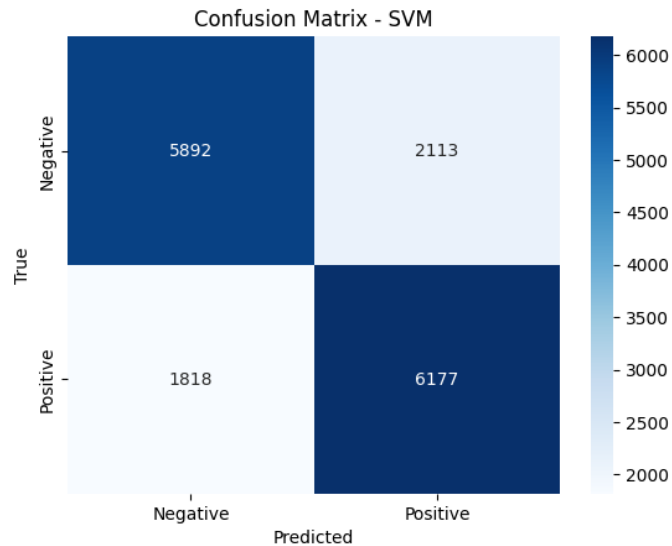
Curva de precisión sensibilidad – MNB



Con el 5% del dataset que son 16000 datos, en el modelo Support Vector Machines (SVM) podemos observar en la siguiente matriz de confusión que existieron 5892 verdaderos negativos y 6177 verdaderos positivos, por otro lado, en los cuadrantes de color blanco vemos que se obtuvo 2113 falsos negativos y 1818 falsos positivos.

Figura 20

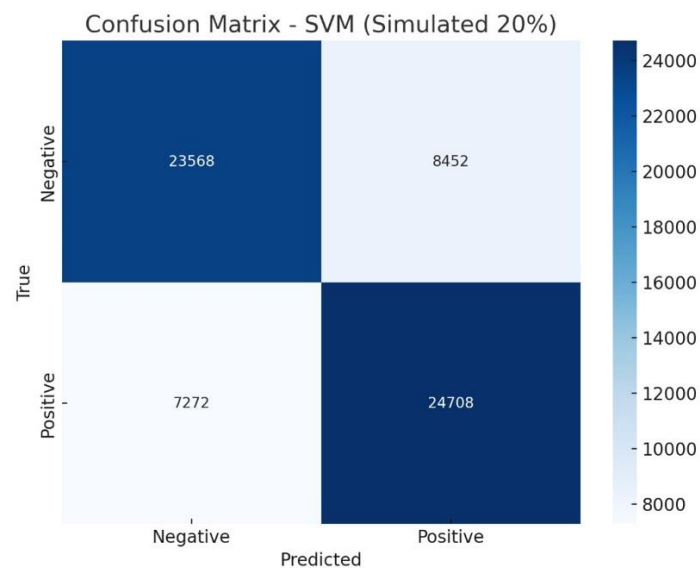
Matriz de confusión – SVM con el 5% dataset



Con el 20% del dataset que son 64000 datos, en el modelo Support Vector Machines (SVM) podemos observar en la siguiente matriz de confusión que existieron 23568 verdaderos negativos y 24708 verdaderos positivos, por otro lado, en los cuadrantes de color blanco vemos que se obtuvo 8452 falsos negativos y 7272 falsos positivos.

Figura 211

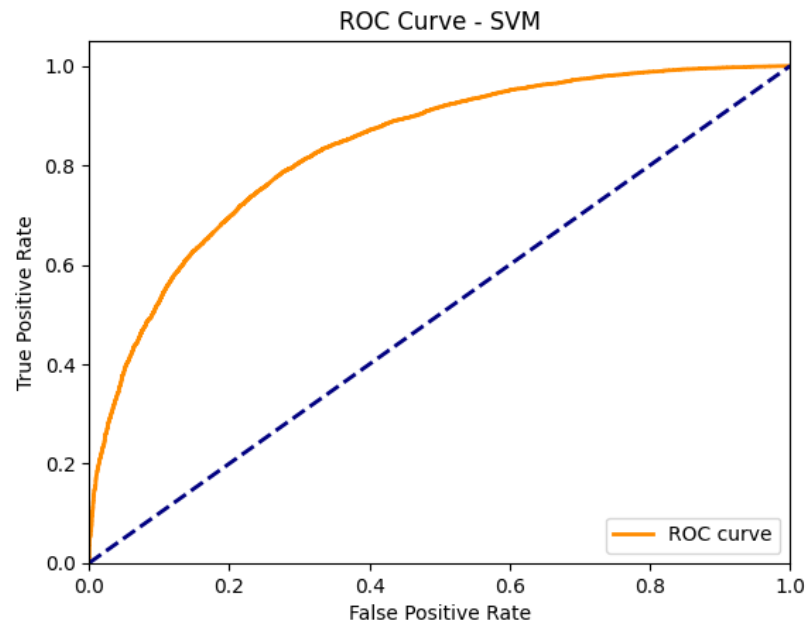
Matriz de confusión – SVM con el 20% dataset



Además, vemos en el modelo Support Vector Machines (SVM) por medio de la curva ROC la Tasa de Verdaderos Positivos representados en el eje y y la Tasa de Falsos Positivos en el eje x.

Figura 222

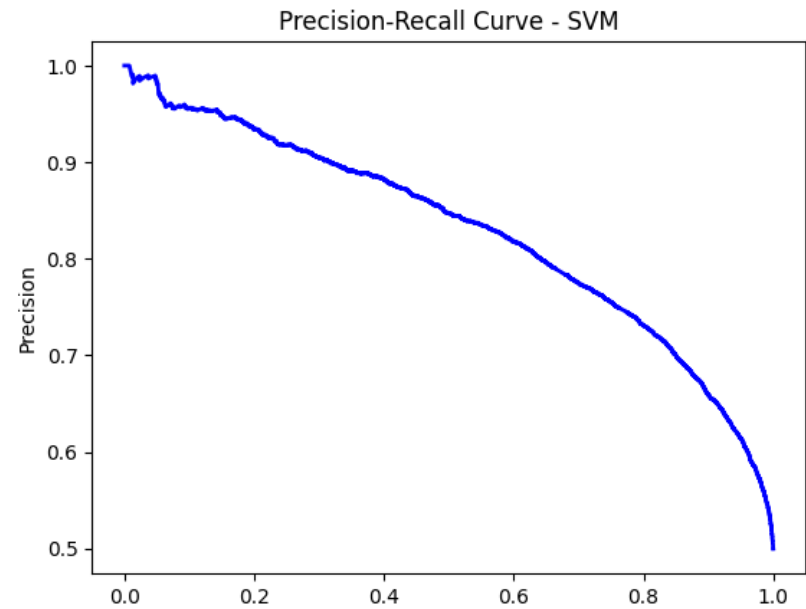
Curva ROC - SVM



Por último, en el modelo Support Vector Machines (SVM) en la curva de precisión-sensibilidad podemos ver el comportamiento de estos dos parámetros, en donde siempre buscaremos tener una alta precisión y una alta recuperación.

Figura 233

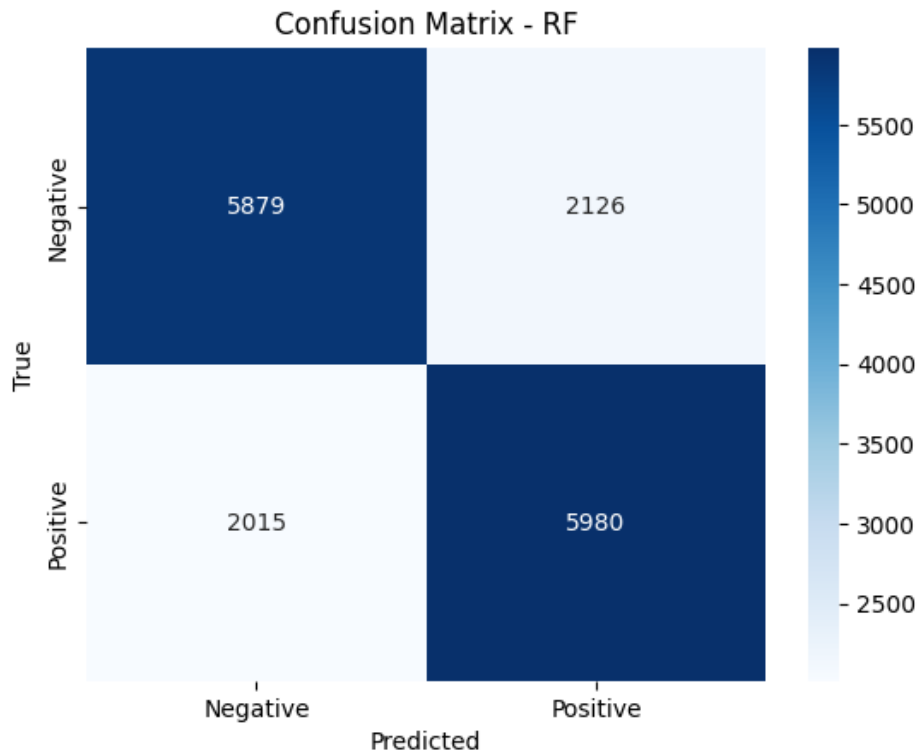
Curva precisión – sensibilidad - SVM



Con el 5% del dataset que son 16000 datos, en el modelo Random Forest podemos observar en la siguiente matriz de confusión que existieron 5879 verdaderos negativos y 5980 verdaderos positivos, por otro lado, en los cuadrantes de color blanco vemos que se obtuvo 2126 falsos negativos y 2015 falsos positivos.

Figura 244

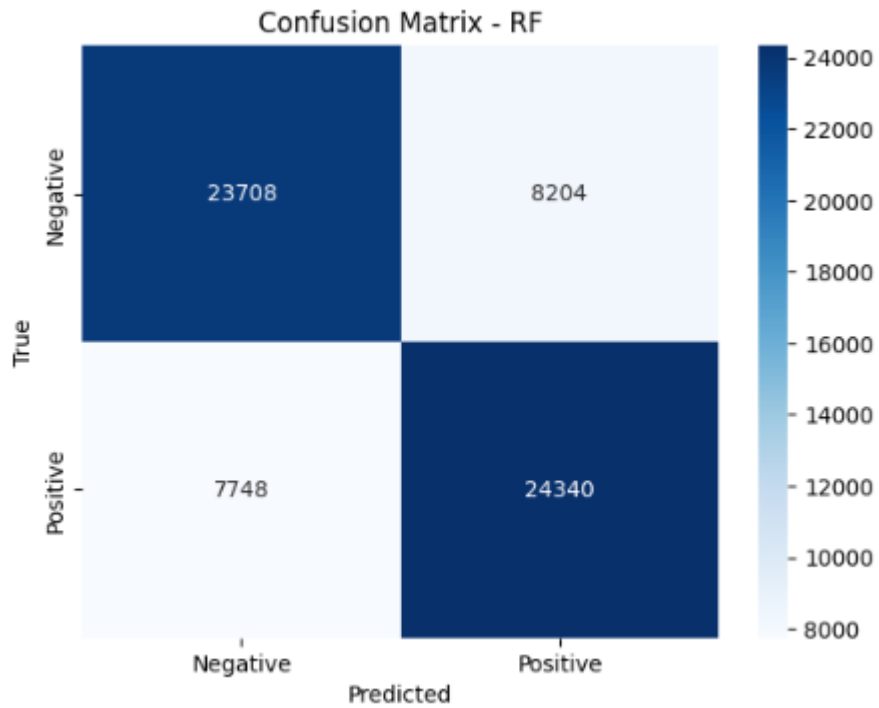
Matriz de confusión RF con el 5% dataset



Con el 20% del dataset que son 64000 datos, en el modelo Random Forest podemos observar en la siguiente matriz de confusión que existieron 5879 verdaderos negativos y 5980 verdaderos positivos, por otro lado, en los cuadrantes de color blanco vemos que se obtuvo 2126 falsos negativos y 2015 falsos positivos.

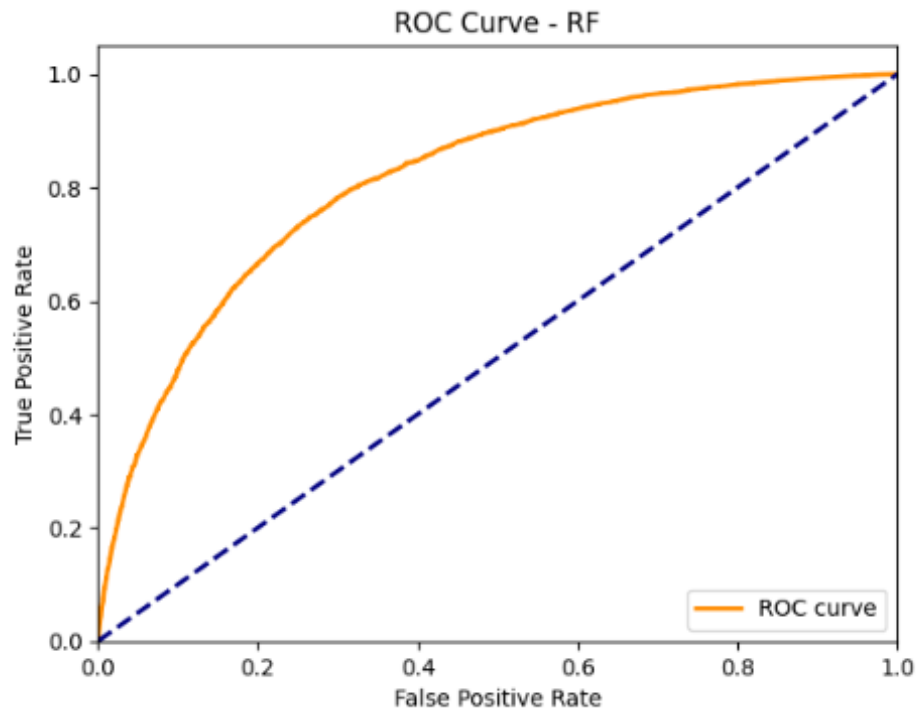
Figura 255

Matriz de confusión RF con el 20% dataset



Además, vemos en el modelo Random Forest por medio de la curva ROC la Tasa de Verdaderos Positivos representados en el eje y y la Tasa de Falsos Positivos en el eje x.

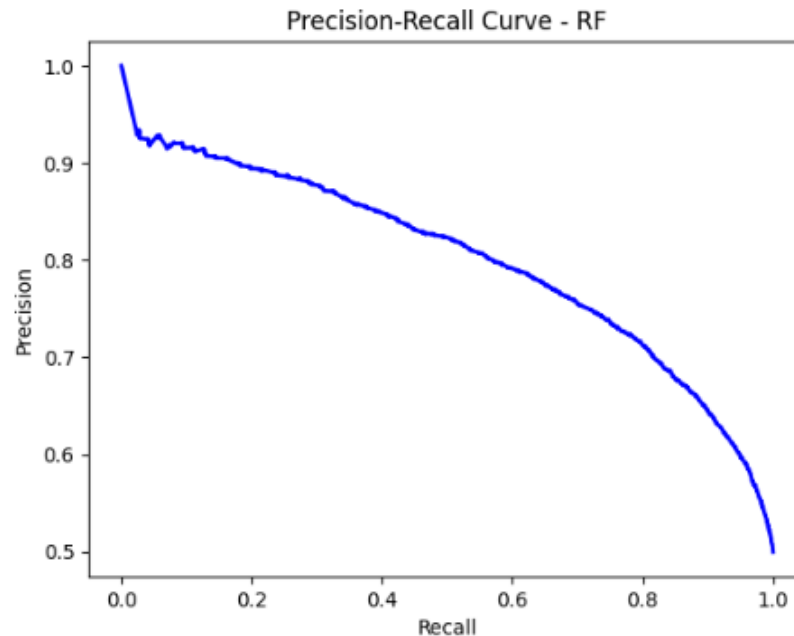
Figura 266

Curva ROC – RF

Por último, en el modelo Random Forest en la curva de precisión-sensibilidad podemos ver el comportamiento de estos dos parámetros, en donde siempre buscaremos tener una alta precisión y una alta sensibilidad.

Figura 277

Curva de precisión – sensibilidad – RF



4.4 Pruebas del sistema

Las pruebas del sistema se realizaron para garantizar que los modelos de análisis de sentimientos se integraban perfectamente en la aplicación final. Para ello, se probaron los modelos con datos nuevos y desconocidos para comprobar su aplicabilidad y solidez. La fase de pruebas incluyó pruebas unitarias para comprobar la funcionalidad de componentes individuales y pruebas de extremo a extremo para validar todo el flujo de trabajo, desde el preprocesamiento de datos hasta la clasificación de sentimientos.

También se realizaron pruebas de rendimiento para evaluar la eficacia y escalabilidad de los modelos. Para ello se midió el tiempo necesario para preprocesar los datos, entrenar los modelos y hacer predicciones. Los resultados confirmaron que los modelos no sólo eran precisos, sino también eficientes, capaces de manejar grandes volúmenes de datos en un plazo razonable.

Estas pruebas garantizaron que el sistema de análisis de sentimientos era fiable, preciso y estaba listo para su despliegue. El sistema se diseñó para manejar entradas de

datos en tiempo real, proporcionando resultados instantáneos de análisis de sentimientos para nuevas publicaciones en redes sociales. Esta exhaustiva fase de pruebas validó la eficacia de los modelos y garantizó que el sistema cumpliera los objetivos del proyecto.

Tabla 5

Tabla resumen de los modelos

	Modelo Naive Bayes	Modelo SVM	Modelo Random Forest
Exactitud	0,7439	0,7543	0,7411
Tiempo preprocesamiento	0,3473 segundos	0,3082 segundos	0,1864 segundos
Tiempo vectorización	0,2934 segundos	0,2537 segundos	0,1558 segundos
Tiempo predicción	0,0023 segundos	40,5680 segundos	2,5370 segundos

Capítulo cinco

Análisis de resultados

5.1 Discusión de resultados

Para tener una idea más clara de la tendencia que tiene el modelo Naive Bayes al incrementar datos para su entrenamiento, a continuación, se muestra las siguientes tablas:

Tabla 6

Tendencia del modelo Naive Bayes

	Verdaderos negativos	Verdaderos positivos	Falsos negativos	Falsos positivos	Total datos
5%	5966	5937	2039	2058	16000
20%	23649	24418	8263	7670	64000

Tabla 7

Tendencia del modelo Naive Bayes

	Exactitud	Precisión	Sensibilidad	Puntuación F1	Soporte
5%	0,7439	0,74	0,74	0,74	16000
20%	0,7510	0,75	0,75	0,75	64000

La tendencia que tiene el modelo Support Vector Machines (SVM) al incrementar datos para su entrenamiento, a continuación, se muestra las siguientes tablas:

Tabla 8

Tendencia del modelo Support Vector Machines (SVM)

	Verdaderos negativos	Verdaderos positivos	Falsos negativos	Falsos positivos	Total datos
5%	5892	6177	2113	1818	16000
20%	23568	24708	8452	7272	64000

Tabla 9

Tendencia del modelo Support Vector Machines (SVM)

	Exactitud	Precisión	Sensibilidad	Puntuación F1	Soporte
5%	0,7543	0,75	0,75	0,75	16000
20%	0,7843	0,78	0,78	0,78	64000

La tendencia que tiene el modelo Random Forest al incrementar datos para su entrenamiento, a continuación, se muestra las siguientes tablas:

Tabla 10

Tendencia del modelo Random Forest (RF)

	Verdaderos negativos	Verdaderos positivos	Falsos negativos	Falsos positivos	Total datos
5%	5879	5980	2126	2015	16000
20%	23708	24340	8204	7748	64000

Tabla 11*Tendencia del modelo Random Forest (RF)*

	Exactitud	Precisión	Sensibilidad	Puntuación F1	Soporte
5%	0,7411	0,74	0,74	0,74	16000
20%	0,7507	0,75	0,75	0,75	64000

5.2 Comparación con trabajos anteriores

De acuerdo a los datos obtenidos se observa que los modelos aplicados de aprendizaje automatizado tienen una exactitud, precisión y sensibilidad por encima del 0,74, indicando que este tipo de aprendizaje nos ayuda a la clasificación de publicaciones en redes sociales; estos datos similares obtuvieron Rojas y Meneses (2022) en su estudio de detección de publicaciones falsas en redes sociales donde los algoritmos en cuanto a precisión fueron sobre los 0,8 y en Random forest también con un gran porcentaje de exactitud, comprobándose que estos modelos son confiables para la implementación de tareas de clasificación debido a su versatilidad, eficacia y la capacidad de manejar grandes volúmenes de datos.

En otro estudio realizado por De la Fuente (2019) sobre la “clasificación de reseña en redes sociales” se aplicó el modelo automatizado con el modelo Naive Bayes, en el cual se observó una exactitud mayor de 0,90 en el análisis de datos binario, y el análisis de datos binarios fue de aproximadamente 0,50; por lo que se especifica que este patrón permite clasificar adecuadamente las reseñas en este caso, es decir, maneja correctamente los datos para la predicción y la clasificación.

Dentro del aprendizaje automatizado en este estudio también se aplicó el modelo Support Vector Machine (SVM) el cual se identifica que tiene una mayor exactitud de los otros dos modelos mencionados, así mismo se aplicó esta prueba en el estudio de Segovia (2022) que señala que este modelo fue el que se utilizó en gran mayoría para realizar detecciones y

predicciones en redes sociales, ya que permite obtener resultados con una alta precisión, permitiendo además dividir los datos en clases que ayudan a tener un mejor rendimiento en los resultados.

Asimismo, un estudio realizado por Kaufhold et al., (2020) sobre la clasificación de publicaciones con mayor relevancia por medio de la aplicación del aprendizaje automatizado, demostró una mejora significativa en la precisión y eficiencia de la clasificación. Esto permitió una gestión más efectiva de la información durante crisis, facilitando respuestas más rápidas por parte de los servicios de emergencia. La conclusión principal destaca la necesidad de adaptar continuamente los algoritmos para manejar dinámicamente la sobrecarga de información en situaciones críticas.

Entre otras investigaciones, realizada por Torres et al., (2020) en el procesamiento de información por medio del aprendizaje automatizado para la clasificación de noticias, donde en la evaluación realizada, el algoritmo de bosques aleatorios demostró ser altamente eficaz, alcanzando una exactitud del 97,22%, precisión del 98,36%, sensibilidad del 98,35%, un puntaje F1 del 98,32%, también destacaron que los algoritmos SVM y MLP arrojaron resultados altos, Además, se reconoció que las fases de recopilación de datos y preprocesamiento son cruciales para el éxito en el desarrollo de sistemas de procesamiento de texto.

Conclusiones

Mediante una investigación sistemática de información se llegó a conocer técnicas de aprendizaje automático, donde se aplicó clasificación binaria o multiclase, colocando etiquetas a las publicaciones de redes sociales como positivas y negativas 1 y 0, respectivamente.

Mediante la aplicación de tres modelos de aprendizaje automatizado como es el modelo de redes bayesianas, modelo de vectores o svm y el modelo de random se obtuvo que la exactitud en los tres modelos aproximadamente fue de 0,74, mientras que en los tiempos de predicción el modelo Bayes tiene un menor tiempo en comparación a SVM y Random.

Por medio de la utilización de datasets que contenían publicaciones de redes sociales, esta información es descargada, procesada y analizada para la posterior clasificación, lo cual se trabajó con bases del 5% y el 20%; identificando que no existen una variación significativa en los resultados, demostrando que los modelos pueden ser aplicables a bases más amplias.

En la evaluación de los modelos se utiliza algunos parámetros de los más conocidos en validación, así como: precisión, exactitud, recall, f1-score, y soporte, en base a los resultados obtenidos el modelo que entrega mejores resultados con los datos de este trabajo es el de vectores o SVM. No obstante, en el tiempo de predicción se ve que el más rápido es el modelo de redes bayesianas.

Recomendaciones

Así como en este trabajo se clasificaron publicaciones como positivas, negativas o neutras, en otras aplicaciones se puede clasificar por puntos geográficos, por fechas y otra característica que sea de algún interés.

Existen otros modelos de aprendizaje automático que también se pueden aplicar en este y otros trabajos, se recomienda la selección de los modelos dependiendo de la aplicación y la dificultad para emplearlos.

Las redes sociales tienen API's propias para desarrolladores, las que permiten tener accesos a los datos de la cuenta propia, en el caso de que se piense realizar una aplicación con el manejo de datos de redes sociales recomiendo hacer uso de estas API's para facilitar el acceso.

Durante este trabajo se observa la importancia de la evaluación de los modelos, ya que así conocemos que tan efectivos son los resultados, recomiendo hacer una evaluación cada vez que se realice el uso de modelos de aprendizaje automático.

Referencias

- Aggarwal, S. (2018). Modern Web-Development using ReactJS. *International Journal of Recent Research Aspects*, 5(1), 2349-7688. Obtenido de <http://ijrra.net/Vol5issue1/IJRRRA-05-01-27.pdf>
- Álvarez, M., Quirós, L., & Cortés, M. (2021). Artificial intelligence and machine learning in medicine. *Food and Drug Administration Journal*, 1-15. Obtenido de <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device#:~:text=Some%20real%2Dworld%20examples%20of,probability%20of%20a%20heart%20attack>.
- Álvarez, N., Pico, P., & Holgado, J. (2020). Detección de Noticias Falsas en Redes Sociales Basada en Aprendizaje Automático y Profundo: Una Breve Revisión Sistemática. *RISTI. Revista Ibérica de Sistemas e Tecnologias de Informação*(1), 1-15. Obtenido de <https://ojs.tdea.edu.co/index.php/cuadernoactiva/article/view/849>
- Blasco, H. (2021). El impacto de las redes sociales en las personas y en la sociedad: redes sociales, redil social, ¿o telaraña? *Revista de Investigación e Innovación educativa Tarbiya*, 49, 1-25. Obtenido de <https://doi.org/10.15366/tarbiya2021.49.007>
- Burdisso, S., Errecalde, M., & Montes, M. (2019). A text classification framework for simple and effective early depression detection over social media streams. *Expert Systems with Applications*, 133(1), 182-197. Obtenido de <https://www.sciencedirect.com/science/article/abs/pii/S0957417419303525>
- De Francisco, A. (2018). *Construcción de modelos de clasificación automática para la detección del acoso*. Madrid: Repoditorio de la Universidad Atónoma de Madrir.

- De la Fuente, Ó. (2019). *Análisis y aplicación de técnicas de aprendizaje automático para clasificación de reseñas en redes sociales*. Leganés: Repositorio Universidad Carlos III de Madrid.
- Díaz, J. (2021). Aprendizaje Automático y Aprendizaje Profundo. *Ingeniare. Revista chilena de ingeniería*, 29(2), 1-3. Obtenido de https://www.scielo.cl/scielo.php?pid=S0718-33052021000200180&script=sci_arttext
- Díaz, J., & Suárez, M. (2019). *Avances y desafíos de métodos y modelos computacionales aplicados al análisis de información en redes sociales* (Vol. 16). España: Actualidad Tecnológica. Obtenido de <https://revistas.udistrital.edu.co/index.php/vinculos/article/view/14714>
- Estévez, J. (2023). *Análisis emocional en redes sociales basados en modelos de aprendizaje automático transformers BERT*. España: Repositorio de la Universidad de Valladolid. Escuela de Ingeniería Informática de Valladolid. Obtenido de <https://uvadoc.uva.es/handle/10324/62911>
- Gómez, E., & Palacios, W. (2021). Revisión de literatura sobre marketing en redes sociales. *Revista Interfaces*, 4(1), 1-21. Obtenido de <https://dialnet.unirioja.es/servlet/articulo?codigo=8661428>
- Gómez, E., Jaimes, R., Hidalgo, O., & Luján, S. (2018). Influencia de redes sociales en el análisis de sentimiento aplicado a la situación política en Ecuador. *Enfoque UTE*, 9(1), 1-12. Obtenido de <http://scielo.senescyt.gob.ec/pdf/enfoqueute/v9n1/1390-6542-enfoqueute-9-01-00067.pdf>
- Guardiola, C. (2020). *Clasificador de texto mediante técnicas de aprendizaje automático*. Valencia: Repositorio Universitat Politècnica de València. Obtenido de <https://riunet.upv.es/bitstream/handle/10251/133840/Guardiola%20-%20Clasificador%20de%20textos%20mediante%20t%c3%a9cnicas%20de%20aprendizaje%20autom%c3%a1tico.pdf?sequence=1&isAllowed=y>

- Isasi, J. (2021). Análisis de sentimientos en R con 'syuzhet'. *The Programming Historian en Español*, 1-6. Obtenido de <https://www.proquest.com/openview/ce524e5e98c861bf1c5430a1637249a8/1?pq-origsite=gscholar&cbl=6458208>
- Kadhim, A. (2019). Encuesta sobre técnicas de aprendizaje automático supervisado para la clasificación automática de textos. *Revista de Inteligencia Artificial*, 1(52), 273-292. Obtenido de <https://link.springer.com/article/10.1007/s10462-018-09677-1>
- Kaufhold, M.-A., Bayer, M., & Reuter, C. (2020). Rapid relevance classification of social media posts in disasters and emergencies: A system and evaluation featuring active, incremental and online learning. *Information Processing and Management*, 57(1), 1-32.
- Leyva, M., Hernández, R., & Estupiñan, J. (2021). Analisis de sentimientos: Una herramienta para estudiar datos cualitativos en la investigación jurídica. *Revista Universidad y Sociedad*, 23(53), 262-266. Obtenido de <https://rus.ucf.edu.cu/index.php/rus/article/view/2476/2432>
- Llerena, P., Proaño, V., Ruíz, T., Ashqui, M., & Chanaluisa, G. (2024). Desarrollo y evaluación de un algoritmo educativo basado en inteligencia artificial para mejorar la enseñanza de la división en estudiantes de secundaria utilizando Python y Google Cola. *Polo del Conocimiento*, 9(1), 1-17. Obtenido de <https://dialnet.unirioja.es/servlet/articulo?codigo=9282012>
- López, J., & Gonzales, F. (2021). Análisis de sentimiento de comentarios en español en Google Play Store usando BERT. *Ingeniare. Revista chilena de ingeniería*, 29(3), 1-11. Obtenido de https://www.scielo.cl/scielo.php?pid=S0718-33052021000300557&script=sci_arttext

- Lovera, F., & Cardinale, Y. (2023). Análisis de sentimientos en Twitter. *RCSI: Revista Científica de Sistemas e Informática*, 3(1), 1-17. Obtenido de <http://209.45.90.234/index.php/rcsi/article/view/418>
- Marín, V., & Cabero, J. (2019). Las redes sociales en educación: desde la innovación a la investigación educativa. *RIED. Revista Iberoamericana de Educación a Distancia*, 22(2), 25-33. Obtenido de <http://revistas.uned.es/index.php/ried/article/view/24248>
- Moreno, Á. (2019). *'Engagement' y redes sociales. Análisis bibliométrico desde el ámbito científico de las relaciones públicas*. Madrid: Repositorio de la Universidad Francisco de Vitoria. Obtenido de <http://ddfv.ufv.es/handle/10641/1934>
- Moreno, A., Pérez, C., & Fernández, C. (2022). Análisis de sentimiento del español basado en corpus. *Taylor and Francis Group an Informa business*, 1-18. Obtenido de <https://www.taylorfrancis.com/chapters/edit/10.4324/9780429329296-38/an%C3%A1lisis-de-sentimiento-del-esp%C3%B1ol-basado-en-corpus-antonio-moreno-ortiz-chantal-p%C3%A9rez-hern%C3%A1ndez-javier-fern%C3%A1ndez-cruz>
- Moreno, F., Martelo, C., Corredor, B., Cifuentes, J., & López, J. (2020). Técnicas para la Clasificación de Sentimientos en Redes Sociales como Apoyo en el Marketing Digital. *Revista Ibérica de Sistemas e Tecnologías de Informação*, 1(35), 167-186. Obtenido de <https://www.proquest.com/openview/10505af13f0740e170a77dec671c59e9/1?pq-origsite=gscholar&cbl=1006393>
- Nguyen, G., Dlugolinsky, S., Bobák, M., Tran, V., López, Á., Heredia, I., . . . Hluchý, L. (2019). Machine Learning and Deep Learning frameworks and libraries for large-scale data mining: a survey. *Artificial Intelligence Review*, 52(1), 77-124. Obtenido de <https://link.springer.com/content/pdf/10.1007/s10462-018-09679-z.pdf>
- Padro, L., & Stanilovsky, E. (2022). *FreeLing 3.0: Towards Wider Multilinguality*. Vataluña: Repositorio de la Universidad de Cataluña.

- Pauli, P. (2019). *Análisis de sentimiento Comparación de algoritmos predictivos y métodos utilizando un lexicon español*. Buenos Aires: Repositorio del ITBA Instituto Tecnológico de Buenos Aires. Obtenido de <https://ri.itba.edu.ar/server/api/core/bitstreams/db2a9097-b8f4-4205-8048-8f9fdc76cd66/content>
- Pérez, S., Martín, A., & Victoria, J. (2019). Evolución de las redes sociales. Las redes sociales como herramienta de marketing digital eficaz para el desarrollo comercial de campañas nacionales. *Revista Inclusiones*, 6(4), 1-27. Obtenido de <https://revistainclusiones.org/pdf31/6%20VOL%206%20NUM%204%20SEVILLA2019OCTUBDICIEMB19INCL.pdf>
- Pianchiche, D. (2023). *Sistema de aprendizaje automático para el análisis de sexismo en el género musical urbano*. Quito: Repositorio PUCE. Obtenido de <https://repositorio.puce.edu.ec/server/api/core/bitstreams/4d097eed-0f9d-4f82-828e-9ecdf5918c48/content>
- Pineda, C. (2022). *Aprendizaje automático y profundo en Python. Una mirada hacia la inteligencia artificial*. Barcelona: RA-MA Editorial. Obtenido de https://books.google.es/books?hl=es&lr=&id=NEi9EAAAQBAJ&oi=fnd&pg=PA7&dq=aprendizaje+autom%C3%A1tico+y+sus+tipos&ots=b3w_7rqbNa&sig=Xns_6L1gB7ZHrs_1t9QXsPPzXoU#v=onepage&q=aprendizaje%20autom%C3%A1tico%20y%20sus%20tipos&f=false
- Quintero, Y., & García, L. (2018). Estudio de análisis de sentimientos basado en aspectos. *IV Conferencia Internacional en Ciencias Computacionales e Informáticas*, 1-10. Obtenido de https://www.researchgate.net/profile/Yisel-Clavel/publication/325370586_A_SURVEY_ON_ASPECT_BASED_SENTIMENT_ANALYSIS/links/5b08704e0f7e9b1ed7f56287/A-SURVEY-ON-ASPECT-BASED-SENTIMENT-ANALYSIS.pdf

- Quirós, L. M., Álvarez, M., & Córtes, M. (2020). Inteligencia artificial y aprendizaje automático en medicina. *Revista Médica Sinergia*, 5(8), 1-12. Obtenido de <https://www.revistamedicasinergia.com/index.php/rms/article/view/557/923>
- Raschka, S., & Mirjalili, V. (2019). Aprendizaje Automático con Python. *Machine Learning Journal*, 1-13. Retrieved from [https://books.google.es/books?hl=es&lr=&id=5EtOEAAAQBAJ&oi=fnd&pg=PT5&dq=Modelos+de+Aprendizaje+Profundo+\(Deep+Learning\)&ots=erFZVv-KK4&sig=ad-AR5Fn0CbftFucAO72_PN--YU](https://books.google.es/books?hl=es&lr=&id=5EtOEAAAQBAJ&oi=fnd&pg=PT5&dq=Modelos+de+Aprendizaje+Profundo+(Deep+Learning)&ots=erFZVv-KK4&sig=ad-AR5Fn0CbftFucAO72_PN--YU)
- Recalde, J. (2021). *El ciberacoso por redes sociales en el Ecuador*. Quito: Repositorio de la Universidad Politécnica Salesiana del Ecuador. Obtenido de <https://dspace.ups.edu.ec/handle/123456789/20945>
- Rodríguez, K., Pardo, M., & Izquierdo, J. (2019). Fundamentación del empleo de las redes sociales educativas en la formación de los profesionales universitarios. *Revista Conrado*, 15(69), 1-6. Obtenido de <http://scielo.sld.cu/pdf/rc/v15n69/1990-8644-rc-15-69-419.pdf>
- Rojas, L., & Meneses, C. (2022). Una comparación empírica de algoritmos de aprendizaje automático versus aprendizaje profundo para la detección de noticias falsas en redes sociales. *Ingeniare. Revista Chilena de Ingeniería*, 30(2), 1-10.
- Ruiz, Z. (2019). *Técnicas de Aprendizaje Automático Aplicadas al Procesamiento de Información Demográfica*. Alicante: Repositorio de la Universidad de Alicante. Obtenido de <https://rua.ua.es/dspace/handle/10045/95608>
- Salamanca, I. (2021). Técnicas de aprendizaje automático aplicadas en los sistemas de predicción. *Revista TIA. Tecnología Investigación y Academia*, 8(1), 1-25. Obtenido de <http://revistas.udistrital.edu.co:8080/index.php/tia/article/view/17325>
- Salazar, C., Aguilar, J., Monsalve, J., & Montoya, E. (2020). Análisis de Sentimientos/Polaridad en diferentes tipos de documentos. *Revista Ibérica de*

Sistemas e Tecnologias de Informação, 1-15. Obtenido de https://www.researchgate.net/profile/Julian-Monsalve-Pulido/publication/350874266_Analisis_de_SentimientosPolaridad_en_diferentes_tipos_de_documentos/links/607748e08ea909241efe0913/Analisis-de-Sentimientos-Polaridad-en-diferentes-tipos-de-documentos.pdf

Salgado , N. (2024). Análisis de sentimientos en datos de redes sociales: aplicación de técnicas de procesamiento de lenguaje natural y machine learning para analizar opiniones y sentimientos en datos de redes sociales en el contexto de sistemas de información. *Revista Científica Dominio de las Ciencias*, 10(1), 314–327. Obtenido de <https://dominiodelasciencias.com/ojs/index.php/es/article/view/3714>

Salina, M., Osio, J., Cappelletti, M., & Morales, M. (2021). Aprendizaje automático aplicado al procesamiento de imágenes para la clasificación de objetos reciclables. *Revista RedUNCI: Red de Universidades con Carreras en Informática*, 31-40. Obtenido de <https://sedici.unlp.edu.ar/handle/10915/130312>

Sánchez , P., Martín, M., & Blanco , D. (2020). Del data-driven al data-feeling: Análisis de sentimientos en tiempo real de mensajes en español sobre divulgación científica usando técnicas de aprendizaje automático. *Anuario Electrónico de Estudios en Comunicación Social "Disertaciones"*, 13(1), 35-58. Obtenido de <https://www.redalyc.org/journal/51115/5111562674006/5111562674006.pdf>

Saura, J., Reyes, A., & Palos, P. (2018). Un Análisis de Sentimiento en Twitter con Machine Learning: Identificando el sentimiento sobre las ofertas de# BlackFriday. *Revista Espacios*, 39(42), 1-15. Obtenido de <https://www.revistaespacios.com/a18v39n42/18394216.html>

Segovia, H. (2022). *Revisión sistemática y análisis de metodologías que utilizan técnicas de minería de datos y aprendizaje automático para detección del trolling en las redes sociales*. Quito: Repositorio Universidad Internacional SEK.

- Syah, M., & Anggraini, D. (2022). Modern Front End Web Architectures with React.Js and Next.Js. *International Research Journal of Advanced Engineering and Science*, 7(1), 132-141. Obtenido de <https://irjaes.com/wp-content/uploads/2022/02/IRJAES-V7N1P162Y22.pdf>
- Tabassum, S., Pereira, F., Fernandes, S., & Gama, J. (2018). Social network analysis: An overview. *Wires data mining and knowledge discovery*, 1-2. Obtenido de <https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/widm.1256>
- Torres, C., Díaz, N., & Cañas, R. (2020). Clasificación de noticias criminales basada en procesamiento del lenguaje natural y algoritmos de aprendizaje automático. *Revista Ibérica de Sistemas e Tecnologías de informação*, 1(38), 117-129.
- Vidal, C., Sánchez, A., Serrano, J., & Rubio, J. (2021). Experiencia académica en desarrollo rápido de sistemas de información web con Python y Django. *Formación universitaria*, 14(5), 1-10. Obtenido de https://www.scielo.cl/scielo.php?pid=S0718-50062021000500085&script=sci_arttext
- Wu, L., & Liu, H. (2018). Tracing Fake-News Footprints: Characterizing Social Media Messages by How They Propagate|. *ACM Digital Library*, 637-645. Obtenido de <https://dl.acm.org/doi/abs/10.1145/3159652.3159677>
- Zaid, M., Rustam, F., Mehmood, A., Mui-zzud-din, Ashraf, I., & Choi, G. (2022). Classification of movie reviews using term frequency-inverse document frequency and optimized machine learning algorithms. *PeerJ Journals*, 1-15. Obtenido de <https://peerj.com/articles/cs-914/>