



DOI: <https://doi.org/10.23857/dc.v9i4.3683>

Ciencias de la Educación
Artículo de Investigación

Análisis comparativo de redes sociales para aplicar técnicas de aprendizaje automático

Comparative analysis of social networks for the application of machine learning techniques

Análise comparativa de redes sociais para aplicação de técnicas de aprendizado de máquina

Freddy Sebastián Rodríguez Alvear ^I
fsrodriguez1@espe.edu.ec
<https://orcid.org/0009-0007-6893-2233>

Joselyn Carolina Rodríguez Alvear ^{II}
jrodriguez1@uce.edu.ec
<https://orcid.org/0000-0002-7494-6099>

Daniel Jossue Arteaga Carbajal ^{II}
darteaga3@indoamerica.edu.ec
<https://orcid.org/0000-0001-8079-088X>

Correspondencia: fsrodriguez1@espe.edu.ec

***Recibido:** 30 de septiembre de 2023 ***Aceptado:** 10 de octubre de 2023 * **Publicado:** 26 de octubre de 2023

- I. Estudiante de la Carrera de Estudiante de la carrera Ingeniería en Tecnologías de la Información (ITIN) de la Universidad de las Fuerzas Armadas – ESPE, Ecuador.
- II. Licenciada en Comunicación Corporativa, Magíster en Dirección de Comunicación Empresarial e Institucional y Magister en Educación Mención en Gestión de Aprendizaje Mediado por TIC, Docente de la Universidad Central del Ecuador.
- III. Abogado de la República del Ecuador, labora en LexCom, Ecuador.

Resumen

Actualmente el análisis de redes sociales se ha convertido en un campo amplio de investigación y para la experimentación se consideró a Twitter como la red social más usada. El objetivo de este trabajo es identificar una red social diferente en la que se pueda aplicar técnicas de aprendizaje automático para detectar hechos y opiniones; de tal forma que se pueda obtener un análisis más amplio de la información que se propaga en las redes sociales. Una de las primeras y más importantes fases dentro del aprendizaje automático es recolectar los datos que luego serán analizados. En este trabajo se aplicó la librería SNScrape con la que se puede recolectar datos y metadatos de varias redes sociales. Cabe mencionar que luego de realizar varias pruebas se concluyó que por la facilidad que presenta el Script aplicado, la eficiencia y tamaño de los datos que se recolectan en LinkedIn es la red que posteriormente será utilizada para llevar a cabo todo el proceso de aprendizaje automático previsto.

Palabras Claves: Desinformación; Información veraz; Difusión de información; Redes sociales; Snscape.

Abstract

The purpose of this study was to determine the incidence of physical activity on emotional intelligence in an educational context. It was developed based on a quantitative research approach of a quasi-experimental design, with a longitudinal explanatory scope. The study sample represented 81 students between 12.4 and 17 years old, with a general average of 14.7, from the sublevels of Higher Basic General Education (EGBS) and Unified General Baccalaureate (BGU), divided into 6 different mixed groups. to whom a different physical activity program was applied. The technique applied was considered to be the survey and as an instrument the “Scale for the evaluation of the expression, management and recognition of emotions” Trait Meta-Mood Scale (TMMS-24). The intervention proposal was directed towards improving the levels of emotional intelligence evidenced in the diagnosis stage. For each subdivided group of the sample, a different type of physical activity program was designed, which addressed the need to include socio-emotional training within the teaching-learning process of Education. The analysis of the results achieved in each state of emotional intelligence in the pre- and post-intervention periods showed positive results in the POST-intervention period, supported by significant differences at a level of $P \leq 0.05$, determining the

Análisis comparativo de redes sociales para aplicar técnicas de aprendizaje automático

effectiveness of all programs. of physical activity proposed. It was concluded that the physical activity programs evaluated have proven to be effective tools to enhance various components of emotional intelligence.

Keywords: Disinformation; truthful information; Dissemination of information; Social networks; Snsrape.

Resumo

Atualmente, a análise de redes sociais tornou-se um amplo campo de pesquisa e para experimentação o Twitter foi considerado a rede social mais utilizada. O objetivo deste trabalho é identificar uma rede social diferenciada na qual técnicas de aprendizado de máquina possam ser aplicadas para detectar fatos e opiniões; de forma que se possa obter uma análise mais ampla da informação que se espalha nas redes sociais. Uma das primeiras e mais importantes fases do aprendizado de máquina é a coleta dos dados que serão posteriormente analisados. Neste trabalho foi aplicada a biblioteca SNScrape com a qual podem ser coletados dados e metadados de diversas redes sociais. Vale ressaltar que após a realização de diversos testes concluiu-se que pela facilidade do Script aplicado, pela eficiência e tamanho dos dados coletados no LinkedIn, é a rede que posteriormente será utilizada para realizar toda a máquina planejada. processo de aprendizado.

Palavras-chave: Desinformação; Desinformação; informações verdadeiras; Disseminação de informação; Redes sociais; Snsrape.

Introducción

En la actualidad las tecnologías de la información y la comunicación (TIC) han tenido un impacto significativo en todas las áreas de actividad humana. Este proceso de digitalización ha acelerado la realización de diversas actividades cotidianas, lo que ha llevado a los ciudadanos a adoptar nuevas estrategias. El aprendizaje en la era digital abarca todos los aspectos de la vida, desde el estudio hasta el trabajo y las actividades de ocio. Gracias al desarrollo de nuevas herramientas de enseñanza y aprendizaje, los estudiantes adquieren conocimientos a través de Internet de manera cada vez más frecuente. Los jóvenes de hoy en día manejan los dispositivos digitales con mayor facilidad que las generaciones anteriores, lo que les permite desarrollar competencias innatas en el uso de las tecnologías (Pegalajar-Palomino y Rodríguez-Torres, 2023; Rodríguez et al., 2023; Urakova et al., 2023).

Análisis comparativo de redes sociales para aplicar técnicas de aprendizaje automático

ResearchGate es considerada una red social digital de investigación, cuyo enfoque principal es desarrollar ciencia de cualquier disciplina; un ejemplo es Twitter, una red social utilizada para publicar información en tiempo real acerca de las temáticas relacionadas en el ámbito político, económico, social, académico, etc. Al igual que Facebook, los usuarios propagan información actual mediante texto, fotos o videos en tiempo real; por su parte Instagram se conecta con la comunidad a partir de fotografías y “reels”, videos cortos; cabe mencionar que la red social que ha causado mayor impacto generacional es Tik Tok, cuyo enfoque principal de esta red social es grabar, editar y compartir videos que se publican en tiempo real con temáticas actuales relacionadas con tecnología, cultura, política, economía y polémica social; finalmente LinkedIn, es considerada como una red social digital para profesionales, que muestra una síntesis del currículum de las personas y se ha convertido en objeto de análisis de muchos investigadores, cuyo enfoque permite definir modelos de propagación de información, otros en mitigar la información que se propaga identificando a los nodos semilla y algunos en el análisis de la desinformación.

Además, en la literatura académica se ha encontrado una clasificación de la información con los términos en inglés: “misinformation”, “disinformation” y “true information”. Sin embargo, no se han encontrado trabajos en los que se identificó tipos de información basados en la argumentación que se propagan intencionalmente en las redes sociales digitales como los hechos y las opiniones utilizando técnicas de aprendizaje automático. La identificación de estos tipos de información, más específica y menos generalizada que la que se ha encontrado, nos permitirá analizar de mejor manera el comportamiento de la información que se propaga en las redes sociales digitales, considerando que en muchas ocasiones puede causar caos y pánico colectivo en la población.

En los trabajos Bharti y Jindal (2020), Rastogi et al. (2021) y Safarnejad et al. (2021) se ha utilizado Twitter como plataforma de experimentación “base” para la investigación; sin embargo, se ha considerado ampliar el análisis de la información buscando veracidad de la misma en relación con otras redes sociales. Por consiguiente, el presente trabajo busca centrarse en realizar pruebas utilizando herramientas de recolección de datos que corresponde a una de las primeras y más importantes fases dentro del proceso de identificación de hechos y opiniones mediante técnicas de aprendizaje automático. Las pruebas fueron realizadas en las redes sociales como: Instagram, Reddit, Facebook y LinkedIn.

Análisis comparativo de redes sociales para aplicar técnicas de aprendizaje automático

Estado del arte Es importante destacar que la propagación de noticias falsas “*fake news*” en las redes sociales puede tener graves consecuencias, ya que puede influir en la toma de decisiones y en la percepción de la realidad de las personas. Starbird (2019) menciona que la propagación de desinformación, información errónea y engaños en las redes sociales puede contribuir a la polarización política, a la creación de burbujas informativas, y puede ser utilizada para influir en elecciones y opiniones públicas. Además, las noticias falsas pueden dañar la reputación de individuos o empresas, y en algunos casos, pueden ser utilizadas para cometer fraudes o generar pánico en la población. Por lo tanto, es importante utilizar herramientas y técnicas efectivas que permitan detectar y prevenir la propagación de noticias falsas en las redes sociales (Starbird et al., 201). El estudio de Vargo y Guo (2018) indicaron que las noticias falsas generaron un impacto significativo en la agenda pública durante el período de estudio. En particular, las noticias falsas sobre la política y los asuntos sociales fueron ampliamente difundidas y tuvieron un mayor impacto en la agenda pública que otras noticias.

Además, el estudio arrojó 40 características distintivas para caracterizar cada NES. Cabe mencionar que, en las características basadas en contenido, se utilizaron consultas lingüísticas y conteo de palabras, así como vectorización de documentos que facilitaron la obtención de 63 y 50 características para cada tweet, respectivamente. Sunstein (2017) examina la plataforma digital Twitter, la cual se ha convertido en un espacio popular para la difusión de opiniones que pueden generar caos en la sociedad debido a la información compartida. En este estudio, se emplea una metodología descriptiva que implica la recolección de tweets para crear un conjunto de características utilizando la técnica TF-IDF, la técnica TF-IDF es una metodología de análisis estadístico que se utiliza comúnmente en minería de datos y procesamiento de lenguaje natural. TF-IDF significa “Term Frequency-Inverse Document Frequency”, Bharti y Jindal (2020) propusieron un modelo de detección automática de rumores en redes sociales mediante la aplicación de la técnica TF-IDF. Según su estudio, esta metodología tiene “el potencial de detectar información falsa rápidamente para contrarrestar la propagación de rumores y generar mayor confianza de los usuarios en las redes sociales” (Bharti & Jindal, 2020, p. 369). La técnica se basa en el conteo de las palabras en un conjunto de documentos y en la identificación de aquellas palabras que son más relevantes para cada documento. La frecuencia de término (TF) mide la cantidad de veces que una palabra específica aparece en un documento, mientras que la frecuencia inversa de documento (IDF) mide la importancia relativa de una palabra en el conjunto de documentos. Al combinar la frecuencia de término y la frecuencia inversa de

Análisis comparativo de redes sociales para aplicar técnicas de aprendizaje automático

documento, TF-IDF puede identificar que las palabras clave más importantes para cada documento en relación con el conjunto completo de documentos. Esta técnica es especialmente útil para el análisis de grandes cantidades de datos no estructurados, como los que se encuentran en las redes sociales, y para la detección de patrones y tendencias en el comportamiento humano y la opinión pública (Bharti y Jindal, 2020). Este mecanismo tiene el potencial de detectar rápidamente información falsa para combatir la propagación de rumores y aumentar la confianza de los usuarios en las redes sociales. La infiltración de información errónea sobre temas de salud como un proceso dinámico de difusión en las redes sociales, incluyendo sus propiedades de contenido. Para investigar este fenómeno, se analizaron las discusiones en Twitter sobre el Zika en 2016, identificando los 264 tweets de desinformación más influyentes y relacionándolos con 455 tweets de información real. Se desarrolló un algoritmo para inferir que la red de propagación de información iterando cada tweet y obteniendo nueve métricas de red. Luego, se utilizó la propagación de información como se la de procesos de Poisson no homogénea (NHPP) y se obtuvieron 40 características distintivas para caracterizar cada NES. Para las características basadas en contenido, se utilizaron consultas lingüísticas y conteo de palabras, así como vectorización de documentos, para obtener 63 y 50 características para cada tweet, respectivamente (Safarnejad et al., 2021).

Ahvanooy et al. (2022) sostienen que la detección de información errónea en las redes sociales continúa siendo un importante desafío debido a la complejidad y dinamicidad de las plataformas en línea, lo que requiere la implementación de nuevos enfoques y herramientas para encontrar que las noticias falsas se distribuyeron de manera abordar este problema. El manejo del volumen, la variedad, velocidad y la precisión de la desinformación en línea, como noticias falsas, publicaciones falsas e imágenes o videos falsos, se ha vuelto cada vez más difícil. En este artículo, se ofrece una visión general de los métodos basados en el aprendizaje automático utilizados para detectar la desinformación y el ocultamiento de información, y se analizan las amenazas y limitaciones actuales de estos enfoques. A partir de estas discusiones, se proponen algunas medidas posibles para contrarrestar estos problemas (Ahvanooy et al., 2022). En su estudio presentado en la Conferencia Internacional de IEEE sobre Salud Digital en julio de 2022, Argyris y colaboradores utilizaron técnicas de aprendizaje profundo para identificar características lingüísticas que pueden afectar la propagación de contenido antivacunas y pro-vacunas en las redes sociales. Las redes sociales han permitido a cualquier persona publicar contenido sin barreras. Aunque esto ha tenido resultados positivos, también ha permitido que los actores malintencionados difundan información errónea entre

Análisis comparativo de redes sociales para aplicar técnicas de aprendizaje automático

los usuarios de la plataforma. Aunque existen filtros y señales útiles para los lectores, como las señales de reputación y las banderas de verificación de hechos, y los algoritmos y moderadores humanos ayudan a detener la difusión de información errónea, ninguna de estas intervenciones aborda la arquitectura básica de las redes sociales que permite la propagación de información errónea.

Las celebridades de las redes sociales, especialmente los usuarios de YouTube han comenzado a participar en los canales principales, como las redes de transmisión y de cable, lo que ha provocado cambios importantes en la forma en que las personas se comportan en las redes sociales (Argyris et al., 2022). En el estudio realizado por Pushkar y Mittal (2022) presentan un análisis del comportamiento del usuario en las redes sociales basado en su interacción en ellas. En la conferencia internacional de Ingeniería y Gestión Inteligente de 2022, los autores presentaron sus investigaciones sobre el análisis de las actividades en las redes sociales y cómo estas pueden ser utilizadas para extraer información valiosa sobre los usuarios y sus intereses. Para Pushkar & Mittal (2022) las actividades y la interacción de los usuarios en las redes sociales pueden ser utilizadas para realizar análisis de comportamiento y cómo estas técnicas pueden ser aplicadas en diversos campos, como la publicidad, el marketing y la investigación de mercado. Rastogi et al. (2021) presentaron un enfoque adaptativo para la detección de noticias falsas en redes sociales, analizando tanto el dominio único como el dominio cruzado. El objetivo del estudio es identificar noticias falsas en redes sociales y evaluar la efectividad del enfoque adaptativo en comparación con otros enfoques. Rastogi et al. (2021) utilizaron un conjunto de datos de noticias falsas de Twitter y realizaron experimentos utilizando una variedad de técnicas de aprendizaje automático, incluyendo clasificadores de árboles de decisión SVM y redes neuronales. Los resultados mostraron que el enfoque adaptativo logró una precisión significativamente mejor en comparación con otros enfoques de detección de noticias falsas Kumari et al. (2021) propusieron un enfoque de aprendizaje multitarea para la detección de noticias falsas que utiliza la novedad, la emoción y el sentimiento para mejorar la precisión del modelo. Por consiguiente, presentaron sus hallazgos en la Conferencia Internacional Conjunta de Redes Neuronales (IJCNN) de 2021. Se ha llevado a cabo una investigación intensiva sobre los modelos automáticos de detección de noticias falsas y verificadores de hechos debido a la reciente explosión de información falsa en las redes sociales. La difusión rápida y peculiar de las noticias falsas y la desinformación han presentado muchos desafíos interesantes a la comunidad de procesamiento de lenguaje natural (PNL) y aprendizaje automático (ML).

Análisis comparativo de redes sociales para aplicar técnicas de aprendizaje automático

La literatura sugiere que la novedad es un elemento clave en la amplificación y vitalidad de la desinformación. Las noticias novedosas y emotivas atraen la atención inmediata del lector, y la emoción está estrechamente relacionada con la presentación de un determinado sentimiento o estado emocional. Kumari et al. (2021) proponen un modelo de aprendizaje multitarea profundo que realiza conjuntamente la detección de novedades, el reconocimiento de emociones, la predicción de sentimientos y la detección de información errónea. Nuestro modelo propuesto ha logrado un rendimiento de última generación (SOTA) para la detección de noticias falsas en tres conjuntos de datos de referencia. En el estudio de Su, Markson y Jones (2019), demuestra como la minería de datos puede utilizarse para analizar grandes cantidades de datos de redes sociales y proporcionar información valiosa para la evaluación de la misma. Su et al. (2019) manifiestan que el Support Vector Machines (SVM) es un modelo de clasificación binaria que tiene la capacidad de identificar y diferenciar entre noticias veraces y falsas. Por esa razón es una de las técnicas más utilizadas. SVM construye un hiperplano en un espacio de alta dimensión para separar diferentes clases de datos y busca el hiperplano óptimo que maximiza la distancia entre los puntos de datos más cercanos de cada clase. SVM es muy efectivo en la clasificación de datos linealmente y no linealmente separables, además, puede manejar datos de alta dimensionalidad y es resistente al sobreajuste (Shu et al, 2017). Sunstein afirma que las técnicas de aprendizaje automático son importantes herramientas para abordar los desafíos que plantea la tecnología digital para la democracia y la diversidad de opiniones. Si se utilizan de manera efectiva, estas técnicas pueden contribuir a crear un entorno en línea más diverso y pluralista, promoviendo una mayor comprensión y cooperación en la sociedad en su conjunto (Sunstein, 2017).

El estudio de Vosoughi, Roy y Aral (2018) se centró en analizar la difusión de noticias verdaderas y falsas en Twitter, y descubrieron que las noticias falsas tienden a difundirse con mayor rapidez y alcance que las noticias verdaderas. El estudio sugiere que esto puede deberse a que las noticias falsas tienden a ser más sensacionalistas y emocionalmente atractivas, lo que las hace más propensas a ser compartidas. La importancia de abordar la difusión de noticias falsas y mejorar la educación del público sobre cómo evaluar la veracidad de la información (Vosoughi et al., 2018). El estudio realizado por Zhang y Varian (2019) se centra en la influencia de Google en el acceso a la información en línea, quienes examinan cómo las personas utilizan Google para acceder a información, cómo se presentan los resultados de búsqueda y cómo esto puede afectar la percepción y la toma de decisiones de los usuarios (Rodríguez-Torres et al., 2018). Zhang & Varían (2019) también discuten las

Análisis comparativo de redes sociales para aplicar técnicas de aprendizaje automático

implicaciones de la concentración del mercado de los motores de búsqueda y la necesidad de una mayor transparencia y regulación. La indagación realizada por Zollo et al. (2017) manifiesta que la propagación de información errónea en línea y cómo la corrección de esta información puede ser más efectiva en algunos contextos que en otros. El estudio encontró que la corrección de información falsa es más efectiva cuando es presentada por fuentes consideradas confiables por los usuarios, y que la pertenencia a grupos o tribus puede influir en la aceptación o rechazo de información corregida. Los autores destacan la importancia de abordar la polarización y el tribalismo en línea para mejorar la corrección de la información errónea (Zollo et al., 2017). En conjunto, los tres artículos subrayan la importancia de abordar el problema de la información errónea en línea y mejorar la capacidad de las personas para evaluar la veracidad de la información. También destacan la necesidad de una mayor transparencia y regulación en la industria de la tecnología y la comunicación en línea.

El objetivo de este trabajo es identificar una red social diferente en la que se pueda aplicar técnicas de aprendizaje automático para detectar hechos y opiniones; de tal forma que se pueda realizar un análisis significativo de la información que se propaga en las redes sociales.

Métodos

Tomando como referencia la revisión de la literatura, se propone una investigación de carácter descriptiva y exploratoria, que busca examinar las redes sociales más usadas y realizar una breve descripción de las características más notables y se determina que la red social más usada es LinkedIn por la forma de interacción entre los usuarios y la facilidad para compartir información en esta plataforma, además en comparación con las otras redes sociales, la comunicación grupal es más frecuente al tener un perfil público, por ende, las publicaciones tendrán mayor alcance (Hernández et al., 2016; Rodríguez, 2015; Rodríguez et al., 2016).

Resultados y discusión

Redes Sociales Las redes sociales son una herramienta de comunicación que posibilita el intercambio de ideas, e información a través de la web, siendo el LinkedIn la red social más utilizada a nivel mundial para un entorno laboral con 774 millones de usuarios y esto se debe a que es una red social es muy atractiva, fácil de usar, interactuar con otros usuarios y permite varios tipos de formatos como video, imagen o texto, tiene perfiles públicos y privados, enlaces con otras redes sociales, es gratuito y tiene seguridad.

Análisis comparativo de redes sociales para aplicar técnicas de aprendizaje automático

Otra de las Redes Sociales más utilizadas es Instagram que brinda más servicios y ha sido capaz de abrirse paso entre los más jóvenes, que la consideran la red social más importante. De hecho, ya cuenta con 1.000 millones de usuarios. Y es muy utilizado entre las generaciones más jóvenes (entre los 16 a 23 años) y por segunda vez consecutiva es una de las redes que más usuarios nuevos atrae. De igual manera Tik Tok es la red social del momento que es una aplicación para crear y compartir videos cortos, como también la red social LinkedIn busca un perfil más profesional. De los sitios analizados ResearchGate permite la difusión de la investigación que brinda más herramientas para que los investigadores encuentren comunidades y exponen los hallazgos a una audiencia más amplia. Permiten subir archivos de publicación, seguir y ser seguidos por comunidades y exponen los hallazgos a una audiencia más amplia. Permiten subir archivos de publicación, seguir y ser seguido por compañeros.

ResearchGate es un sistema de puntos que combina descargas de publicaciones, citas, participación en foros de discusión y número de visitas. Además, ResearchGate permite vincular las cuentas de redes sociales no académicas de los usuarios, como Twitter y Facebook.

Estándares de elección

Los estándares de elección se derivaron de la exploración de la documentación existente y se fundamentaron en las siguientes particularidades: (1) Foros de discusión, (2) Repositorio, (3) Comunicaciones, (4) Referencias bibliográficas, (5) Métricas alternativas, (6) Perfiles, (7) Comunicación grupal, (8) Visibilidad en la red, (9) Carga de publicaciones, (10) Enlaces de redes sociales, (11) Seguridad, (12) Gratuidad.

Las plataformas y herramientas en línea para la investigación académica se pueden clasificar en diferentes categorías según su función y características. (ver Tabla. 1). A continuación, se describen algunas de estas categorías:

- **Foros de discusión:** son espacios en línea donde los investigadores pueden discutir y debatir sobre temas específicos relacionados con sus campos de investigación. Estos foros suelen estar moderados por expertos y pueden ser públicos o privados.
- **Repositorio:** son plataformas que almacenan y difunden publicaciones académicas, como artículos, tesis, informes y otros tipos de documentos. Estos repositorios pueden ser institucionales, temáticos o de acceso abierto.

Análisis comparativo de redes sociales para aplicar técnicas de aprendizaje automático

- **Comunicaciones:** son herramientas que permiten a los investigadores comunicarse y colaborar en línea, ya sea a través de mensajes, correos electrónicos, videoconferencias o chats.
- **Referencias bibliográficas:** son herramientas que ayudan a los investigadores a gestionar y citar fuentes bibliográficas de manera eficiente. Estas herramientas pueden ser gratuitas o de pago y ofrecen diferentes funciones, como importación de referencias, organización de bibliografías y generación de citas en diferentes estilos.
- **Métricas alternativas:** son herramientas que miden el impacto de las publicaciones académicas y las investigaciones en línea, utilizando diferentes criterios y fuentes de datos. Estas métricas pueden incluir el número de menciones en redes sociales, el número de descargas o el número de citas en fuentes no tradicionales.
- **Perfiles:** son herramientas que permiten a los investigadores crear y mantener perfiles en línea que contienen información sobre su trabajo, publicaciones, colaboraciones y otros aspectos relevantes de su carrera académica.
- **Comunicación grupal:** son herramientas que permiten a los investigadores colaborar y comunicarse en grupo, como por ejemplo en proyectos de investigación, tareas, reuniones y otros eventos.
- **Visibilidad en la red:** son herramientas que ayudan a los investigadores a mejorar su visibilidad en línea, ya sea a través de la optimización de motores de búsqueda, la promoción en redes sociales o la participación en eventos y grupos en línea.
- **Carga de publicaciones:** son herramientas que permiten a los investigadores cargar y compartir sus publicaciones académicas en línea ya sea a través de repositorios institucionales, redes sociales académicas o plataformas especializadas.
- **Enlaces de redes sociales:** son herramientas que permiten a los investigadores compartir y promocionar sus publicaciones y trabajos en redes sociales como Twitter, LinkedIn y Facebook.
- **Seguridad:** son herramientas que garantizan la seguridad y privacidad de los datos de los investigadores, incluyendo el cifrado de datos, la autenticación de usuarios y la protección contra ataques cibernéticos.

Análisis comparativo de redes sociales para aplicar técnicas de aprendizaje automático

- **Gratuidad:** son herramientas y plataformas que ofrecen servicios gratuitos para los investigadores, ya sea para el acceso a publicaciones, la carga de publicaciones o la promoción de trabajos en línea.

Biblioteca beautifulsoup4

BeautifulSoup4 es una biblioteca de Python que se utiliza para analizar documentos HTML y XML, permite buscar, navegar, modificar y extraer información de estos documentos de forma fácil y rápida. Algunas de las características más destacadas de BeautifulSoup4 son: (1) Soporte para analizar documentos HTML y XML (2) Capacidad para trabajar con documentos imperfectos o malformados (3) Permite buscar elementos HTML o XML por nombre de etiqueta, atributos, texto o contenido (4) Proporciona funciones para navegar por la estructura del (5) Documento, como acceder a los elementos padres, hijos y hermanos (6) Permite modificar documentos y crear nuevos elementos HTML o XML

Bibliotecas pandas

Es una herramienta muy popular y ampliamente utilizada en el mundo de la ciencia de datos y la programación en Python. Pandas proporciona estructuras de datos de alto nivel y herramientas para manipular y analizar datos de diversas fuentes, incluyendo archivos CSV, bases de datos SQL y hojas de cálculo de Excel. Algunas de las características más importantes de pandas son:

- (1) Estructuras de datos flexibles y eficientes para trabajar con datos tabulares, como Series (unidimensionales) y Data-Frames (bidimensionales).
- (2) Funciones para la manipulación de datos, como la limpieza y la transformación de datos, la selección y filtrado de datos, y la agregación de datos.
- (3) Integración con otras herramientas de análisis de datos de Python, como NumPy, matplotlib y scikit-learn.
- (4) Soporte para el procesamiento de grandes conjuntos de datos y la integración con otras herramientas de Big Data como Apache Spark.

Tabla 1

Análisis comparativo de las redes sociales

Análisis comparativo de redes sociales para aplicar técnicas de aprendizaje automático

CRITERIOS	FACEBOOK	INSTAGRAM	TWITTER	LINKEDIN	TELEGRAM	WHATSAPP	TIK TOK	ResearchGate
Foros de discusión	☑	☑	☑	☑	☑	☑	☑	✗
Repositorio	✗	✗	✗	✗	✗	✗	✗	✗
Comunicaciones	☑	☑	☑	☑	☑	☑	☑	☑
Referencias bibliográficas	✗	✗	✗	✗	✗	✗	✗	✗
Métricas alternativas	✗	✗	✗	✗	✗	✗	✗	☑
Perfiles Públicos - Privados	✗	☑	✗	☑	✗	✗	☑	☑
Comunicación grupal	☑	✗	✗	☑	☑	☑	✗	✗
Visibilidad en la red	☑	☑	☑	☑	☑	☑	☑	☑
Carga de publicaciones	☑	☑	☑	☑	✗	✗	☑	☑
Enlaces a redes sociales	☑	☑	☑	☑	☑	☑	☑	☑
Seguridad	☑	☑	☑	☑	☑	☑	☑	☑
Gratis	☑	☑	☑	☑	☑	☑	☑	☑
TOTAL	8	8	7	9	7	7	8	8

Implementación

La implementación de una script en LinkedIn requiere del uso de diversas herramientas y conocimientos para acceder a los datos y funcionalidades que ofrece la plataforma. Entre estas herramientas, Python es una de las más utilizadas para la minería de datos en LinkedIn, permitiendo extraer información valiosa de los perfiles de los usuarios, como su experiencia laboral, habilidades, educación, contactos, entre otros. Para utilizar LinkedIn con Python, es necesario tener una cuenta de desarrollador y generar las credenciales correspondientes. Una vez obtenidas las credenciales, se pueden utilizar bibliotecas de Python, como python-linkedin, para conectarse a LinkedIn y extraer la información deseada. La minería de datos en LinkedIn puede ser útil para diferentes propósitos, como la búsqueda de empleo, la identificación de perfiles de interés, el análisis de tendencias y la creación de redes de contactos. Esta práctica es cada vez más común entre los reclutadores y empresas que buscan talento en la red social, ya que permite acceder a información valiosa sobre los candidatos y realizar una selección más precisa y efectiva.

Recopilar datos de LinkedIn puede ser importante por varias razones, como la identificación de clientes potenciales, el conocimiento de los competidores, la contratación de talentos y el análisis de tendencias en diferentes industrias y sectores. La recopilación y análisis de datos en LinkedIn puede proporcionar información valiosa que permita tomar decisiones informadas y mejorar estrategias comerciales. Es importante destacar que la extracción de información personal de empleados de una empresa sin su consentimiento es ilegal y va en contra de la ética y la privacidad de los individuos. Por lo tanto, cualquier uso de herramientas y técnicas para acceder a dicha información debe ser realizado de manera ética y legal, respetando siempre la privacidad y los derechos de los empleados. En caso de que se requiera acceder a información de los empleados de una empresa para fines

Entre las ventajas de los scripts para los investigadores se incluyen:

- **Acceso a datos estructurados y no estructurados:** los scripts permiten acceder a diferentes tipos de datos, como texto, imágenes, audio, video, redes sociales y datos científicos, entre otros.
- **Integración de datos:** los scripts permiten combinar y correlacionar datos de diferentes fuentes para obtener nuevos conocimientos y descubrir patrones ocultos.
- **Automatización de tareas:** los scripts permiten automatizar tareas repetitivas, como la descarga y el procesamiento de datos, lo que reduce el tiempo y el esfuerzo requeridos para llevar a cabo investigaciones.
- **Análisis de grandes cantidades de datos:** los scripts permiten procesar grandes cantidades de datos de manera rápida y eficiente, lo que permite a los investigadores realizar análisis y experimentos a gran escala.

Conclusiones

Se puede concluir que, la implementación de un scripts en LinkedIn enfocado a la detección de hechos y opiniones puede ser de gran utilidad para investigadores y empresas que buscan recopilar información sobre las tendencias y opiniones de los usuarios de la plataforma. Los scripts de LinkedIn permiten acceder a grandes cantidades de datos y analizarlos para extraer información relevante, como las tendencias en la opinión pública sobre temas específicos o el sentimiento de los usuarios hacia una marca o producto en particular. Además, los scripts de LinkedIn pueden ayudar a identificar y monitorear noticias falsas y desinformación en la plataforma, lo que es particularmente importante en el contexto de la difusión de información errónea en línea.

Sin embargo, es importante tener en cuenta que la detección de hechos y opiniones a través de scripts puede ser un proceso complejo y requerir el uso de técnicas de procesamiento del lenguaje natural y aprendizaje automático avanzadas. Además, el análisis de datos a través de scripts también plantea desafíos en términos de privacidad y ética, ya que se debe garantizar que los datos se manejen de manera responsable y se respeten los derechos de los usuarios de la plataforma. Por lo tanto, es importante que los investigadores y empresas que utilizan scripts en LinkedIn para la detección de hechos y opiniones se adhieran a las políticas de privacidad y ética establecidas por la plataforma y utilicen estas herramientas de manera responsable y transparente.

Referencias

- Argyris, Y. A., Zhang, N., Bashyal, B., & Tan, P. N. (2022, July). Using Deep Learning to Identify Linguistic Features that Facilitate or Inhibit the Propagation of Anti-and Pro-Vaccine Content on Social Media. In 2022 IEEE International Conference on Digital Health (ICDH) (pp. 107-116). IEEE.
- Ahvanooey, M. T., Zhu, M. X., Mazurczyk, W., Choo, K. K. R., Conti, M., & Zhang, J. (2022). Misinformation Detection on Social Media Challenges and the Road Ahead. *IT Professional*, 24(1), 34-40.
- Bharti, M., & Jindal, H. (2020, November). Automatic rumour detection model on social media. In 2020 Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC) (pp. 367-371). IEEE.
- Hernández, R., Fernández, C., y Baptista, P. (2016). *Metodología de la Investigación*. MC Graw Hill Education.
- Kumari, R., Ashok, N., Ghosal, T., & Ekbal, A. (2021, July). A multitask learning approach for fake news detection: novelty, emotion, and sentiment lend a helping hand. In 2021 International Joint Conference on Neural Networks (IJCNN) (pp. 1-8). IEEE.
- Pegalajar-Palomino, M. D. C. y Rodríguez-Torres, A. F. (2023). Las competencias digitales en estudiantes de las carreras de Educación en Ecuador. *Campus Virtuales*, 12(2), 113-126. <https://doi.org/10.54988/cv.2023.2.1215>
- Pushkar, P., & Mittal, U. (2022, April). User Behavior Analysis based on their Social Media Interaction. In 2022 3rd International Conference on Intelligent Engineering and Management (ICIEM) (pp. 107-110). IEEE.
- Rastogi, S., Gill, S. S., & Bansal, D. (2021, December). An Adaptive Approach for Fake News Detection in Social Media: Single vs Cross Domain. In 2021 International Conference on Computational Science and Computational Intelligence (CSCI) (pp. 1401-1405). IEEE.
- Rodríguez, Á., Cargua, A., Cargua, N., y Garcés, J. (2023). Competencias Digitales de los Estudiantes de la Carrera de Pedagogía de la Actividad Física y Deporte: Caso Ecuatoriano. En A. Morales, A. Vargas, J., J. Martínez-Iglesias y C. Gallardo (Coords.). *Innovación y Transferencias de Conocimientos*. (pp. 81-96). Dynkinson, S.L.

Análisis comparativo de redes sociales para aplicar técnicas de aprendizaje automático

- Rodríguez, Á., Gómez, M., Granda, V., y Naranjo, J. (2016). Paradigmas de investigación: tres visiones diferentes de ver y comprender a la Educación Física. *Lecturas: Educación Física y Deportes. Revista Digital*, 21(222), 1-12. <https://doi.org/10.46642/efd.v26i275.2819>
- Rodríguez, A. (2015). La formación inicial y permanente de los docentes de enseñanza no universitaria del Distrito Metropolitano de Quito y su influencia en los procesos de enseñanza y aprendizaje, la evaluación institucional, el funcionamiento, la innovación y la mejora de los centros educativos. Tesis Doctoral Inédita. San Sebastián (España). Universidad del País Vasco. <https://doi.org/10.2307/j.ctv253f6hj.59>
- Rodríguez-Torres, Á., Rosero-Duque, M., & Aguirre-Obando, E. (2018). La búsqueda de la información científica en la Universidad Central del Ecuador: Reflexiones desde el caso Facultad de Cultura Física. *Revista Ciencias Sociales*, 1(39), 181–188. https://doi.org/10.29166/ciencias_medicas.v43i1.1462
- Safarnejad, L., Xu, Q., Ge, Y. & Chen, S. (2021). A Multiple Feature Category Data Mining and Machine Learning Approach to Characterize and Detect Health Misinformation on Social Media. *IEEE Internet Computing*, 25(5), 43-51.
- Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explorations Newsletter*, 19(1), 22-36.
- Starbird, K., Arif, A., Wilson, T., & Stanek, S. (2019). Disinformation, misinformation, and hoax cascades in social media: Mapping content propagation. *Annual Review of Information Science and Technology*, 53(1), 489-516.
- Su, Y., Markson, C., & Jones, K. T. (2019). The Big Five personality traits and ideological attitudes: An empirical study using Facebook data. *Personality and Individual Differences*, 142, 17-22.
- Sunstein, C. R. (2017). *Republic: Divided democracy in the age of social media*. Princeton University Press
- Vargo, C. J., & Guo, L. (2018). The agenda-setting power of fake news: A big data analysis of the online media landscape from 2014 to 2016. *New Media Society*, 20(5), 2028-2049.
- Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146-1151.
- Urakova, F. K., Ishmuradova, I. I., Kondakchian, N. A., Akhmadieva, R. S., Torkunova, J. V., Meshkova, I. N., & Mashkin, N. A. (2023). Investigating digital skills among Russian higher

Análisis comparativo de redes sociales para aplicar técnicas de aprendizaje automático

education students. *Contemporary Educational Technology*, 15(1), ep398, 1-13.
<https://doi.org/10.30935/cedtech/12600>

Zhang, Y., & Varian, H. (2019). Understanding the role of Google in information access. *Journal of Economic Perspectives*, 33(2), 197-223

Zollo, F., Bessi, A., Del Vicario, M., Scala, A., Caldarelli, G., Shekht- man, L., ... Quattrociocchi, W. (2017). Debunking in a world of tribes. *PLoS One*, 12(7), e0181821.

©2023 por los autores. Este artículo es de acceso abierto y distribuido según los términos y condiciones de la licencia Creative Commons Atribución-NoComercial-CompartirIgual 4.0 Internacional (CC BY-NC-SA 4.0) (<https://creativecommons.org/licenses/by-nc-sa/4.0/>).