

DOI: <https://doi.org/10.46296/ig.v8i15.0231>

ANÁLISIS DE MODELOS PREDICTIVOS EN INSTAGRAM: TENDENCIAS ACTUALES Y BRECHAS EN LA INVESTIGACIÓN

PREDICTIVE MODEL ANALYSIS ON INSTAGRAM: CURRENT TRENDS AND RESEARCH GAPS

Loor-Bravo Brigitte Nathaly ¹; Rivadeneira-Barreiro Lucía Bernarda ²

¹ Universidad Técnica de Manabí. Portoviejo, Ecuador.

Correo: bloor6116@utm.edu.ec. ORCID ID: <https://orcid.org/0009-0005-5075-5609>.

² Universidad Técnica de Manabí. Portoviejo, Ecuador.

Correo: lucia.rivadeneira@utm.edu.ec. ORCID ID: <https://orcid.org/0000-0001-5989-7703>.

Resumen

Las predicciones en redes sociales permitieron comprender el comportamiento de los usuarios e identificar tendencias, optimizando las estrategias de contenido. A pesar de investigaciones previas en diversas plataformas, persistió una brecha en el análisis de modelos predictivos enfocados en Instagram. Por ello, este estudio tuvo como propósito realizar una revisión sistemática de la literatura en bases de datos académicas sobre modelos de predicción de popularidad en esta red social. Los resultados revelaron que las organizaciones emplearon estos modelos para mejorar decisiones de marketing, detectar tendencias y personalizar contenido. Se encontró que las publicaciones con fotos individuales fueron las más investigadas, mientras que formatos como reels y carruseles plantearon desafíos analíticos debido a la necesidad de técnicas avanzadas de preprocesamiento de texto para mejorar la precisión. Aunque se avanzó en el desarrollo de modelos para otras redes, los estudios específicos en Instagram presentaron limitaciones en cuanto a transparencia, ética y privacidad, además de un uso reducido de ciertos tipos de contenido para realizar predicciones. Las implicaciones de este estudio son prácticas, ya que proporcionan a los profesionales de marketing y a los creadores de contenido una comprensión integral de cómo los modelos predictivos en Instagram pueden optimizar la estrategia de contenido en esta plataforma.

Palabras clave: modelo predictivo, Instagram, redes sociales, tendencias, brechas.

Abstract

Social media predictions allowed for an understanding of user behavior and the identification of trends, optimizing content strategies. Despite prior research across various platforms, a gap remained in the analysis of predictive models focused on Instagram. Therefore, this study aimed to conduct a systematic literature review in academic databases on models for predicting popularity on this social network. The findings revealed that organizations employed these models to improve marketing decisions, detect trends, and personalize content. It was found that individual photo posts were the most studied, while formats like reels and carousels posed analytical challenges due to the need for advanced text preprocessing techniques to enhance accuracy. Although progress has been made in developing models for other networks, studies specific to Instagram still presented limitations regarding transparency, ethics, and privacy, as well as a limited use of certain types of content for making predictions. The implications of this study are practical, as they provide marketing professionals and content creators with a comprehensive understanding of how predictive models on Instagram can be used to optimize content strategy on this platform.

Keywords: predictive model, Instagram, social media, trends, gaps.

Información del manuscrito:

Fecha de recepción: 03 de octubre de 2024.

Fecha de aceptación: 19 de diciembre de 2024.

Fecha de publicación: 10 de enero de 2025.



1. Introducción

Actualmente, las redes sociales se han convertido en un insumo importante dentro del campo de la ciencia de datos, comunicación y el marketing. Esto se debe a que la generación y consumo de información en redes sociales se ha popularizado durante los últimos años, haciendo que estas plataformas sean una fuente relevante de contenido que permite un mayor entendimiento de las interacciones de los usuarios. Entre las diversas plataformas que han surgido en la última década, Instagram se ha posicionado como la red social más popular (Casaló et al., 2021; Saeidi & Baradari, 2023), con cerca de 1.4 billones de usuarios al 2024 (Dixon, 2024), a causa de su atractivo visual, interfaz amigable para los usuarios, al sentido de comunidad que se genera mediante el uso de hashtags, comentarios o mensajes directos y a la necesidad de gratificación (Diefenbach & Anders, 2022). Debido a esta popularidad, Instagram se ha convertido en un espacio clave para organizaciones, usuarios y creadores de contenido que buscan captar la atención de su público.

Dentro de este contexto, la predicción de la popularidad de las publicaciones en Instagram ha atraído el interés tanto de académicos como de profesionales dentro de las organizaciones. La importancia de entender e implementar predicciones en esta red social radica en la capacidad para anticipar el comportamiento de los usuarios, permitiendo a los creadores de contenido adaptar sus estrategias para mejorar la visibilidad y el alcance de sus publicaciones (Chaudhary et al., 2021). Además, contar con modelos predictivos precisos puede significar la diferencia entre el éxito y el fracaso dentro de campañas de marketing (Yu & Egger, 2021). Los modelos predictivos permiten entender cuál es el mejor momento para publicar, los tipos de contenido preferidos entre los usuarios, la mitigación de riesgos y determinar las estrategias que pueden incrementar la participación e interacción de los usuarios (Ibrahim & Aljarah, 2023). Los modelos predictivos optimizan también la planificación y ejecución de campañas en Instagram, así como la elaboración de estrategias de marketing más efectivas y basadas en datos, lo cual permite

una mayor eficacia en la generación de contenido.

El desarrollo e implementación de modelos predictivos se ha basado mayormente en algoritmos de aprendizaje automático, conocido en inglés como machine learning. Cuando los algoritmos son supervisados se desarrollan las predicciones usando datos históricos, los cuales se dividen en subconjuntos de datos de entrenamiento y de prueba. En el entrenamiento, los algoritmos aprenden de los patrones y relaciones encontrados entre las variables independientes, que son las predictoras, y su influencia en la variable dependiente, que es la resultante. Este conocimiento es evaluado en el subconjunto de prueba, que contiene datos que el algoritmo no ha observado previamente. Los algoritmos de supervisión, como las redes neuronales o máquinas de soporte de vectores (SVM) son versátiles ya que pueden aplicarse tanto a problemas de clasificación como de regresión (Mohammed & Kora, 2023). Estos algoritmos aprenden de patrones complejos en datos, independientemente de su forma o

estructura. En tareas de clasificación, pueden identificar y predecir categorías discretas, como en el reconocimiento de imágenes o la categorización de texto. En contraste, en tareas de regresión, pueden predecir valores continuos, como el precio de una vivienda. Además, los modelos entrenados en una tarea pueden transferir conocimientos a otras tareas relacionadas, lo que permite el desarrollo de soluciones integrales que combinan ambas capacidades en aplicaciones del mundo real. Sin embargo, la efectividad de estos modelos también depende de la calidad y la cantidad de datos disponibles, así como de la capacidad de los algoritmos para manejar la variabilidad inherente en las interacciones de los usuarios.

En Instagram, investigaciones previas se han enfocado en modelos predictivos en distintas áreas de conocimiento. Por ejemplo, en el campo de turismo para predecir cómo los colores de las fotografías influyen en el número de me gusta y comentarios (Yu & Egger, 2021), en el marketing para reforzar las relaciones entre clientes y marcas (Ibrahim & Aljarah, 2023), dentro de

áreas médicas para predecir depresión (Chiu et al., 2021) o para aspectos relacionados a la personalidad (Wu et al., 2022). A pesar de estos avances en la literatura, y hasta el mejor de nuestro conocimiento, no existen estudios previos que abarquen de forma integral los modelos predictivos específicamente diseñados para Instagram, considerando carruseles o carretes, que es un conjunto de fotos o videos en una misma publicación, y reels que son videos cortos. Debido a esta necesidad, este estudio tiene como propósito realizar una revisión de literatura sobre las tendencias actuales de modelos en Instagram e identificar las brechas en el conocimiento con relación a los temas donde aún es necesario desarrollar nuevas investigaciones. Para ello, la metodología utilizada incluye una búsqueda exhaustiva en las bases de datos Scopus, Web of Science, Taylor & Francis, IEEE y Springer, utilizando palabras clave en inglés y español relacionadas con modelos predictivos en Instagram. Los resultados de esta revisión indican que, aunque ha habido un progreso en el desarrollo de modelos predictivos en Instagram, aún hay

brechas en la literatura relacionados con la transparencia y explicabilidad de los algoritmos, consideraciones éticas y privacidad, así como modelos predictivos que usen carruseles y reels como fuentes de datos.

El resto del documento se estructura de la siguiente forma: En la Sección 2 se explica la metodología del estudio, donde se detalla cómo se recolectaron y analizaron los datos para la investigación. Posteriormente, se presenta en la Sección 3 los resultados y la discusión, donde se realiza un análisis crítico del estado del arte de modelos predictivos en Instagram, así como la identificación de las brechas del conocimiento. Finalmente, en la Sección 4 se presentan las conclusiones principales del estudio, limitaciones y áreas para investigaciones futuras.

2. Materiales y métodos

Es sección expone los pasos que se realizó para determinar las tendencias dentro del área de conocimiento y brechas siguiendo la metodología PRISMA (Page et al., 2021). PRISMA proporciona un

marco estructurado para desarrollar una revisión de literatura transparente, rigurosa y reproducible por otros investigadores. Como parte del enfoque de esta investigación se formularon las siguientes preguntas de investigación (PI) que se orientan al desarrollo del estudio:

PI1: ¿Cómo utilizan las organizaciones los modelos predictivos basados en datos de Instagram?

PI2: ¿Qué tipos de publicaciones y métricas son más efectivas en los modelos predictivos de popularidad

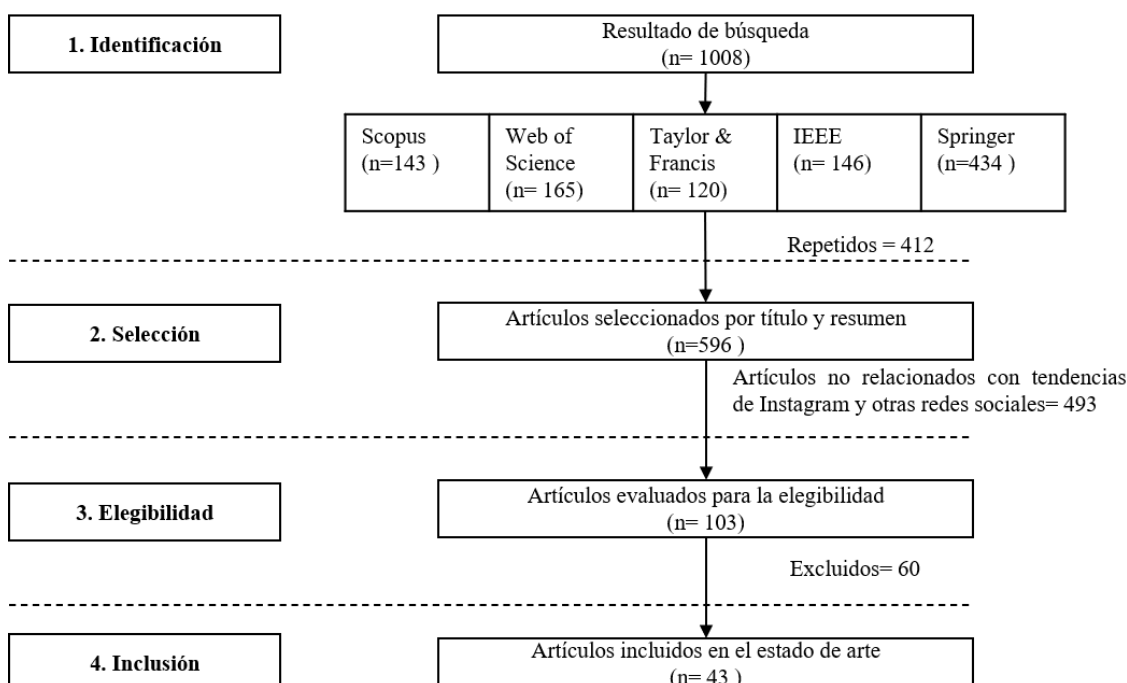
en Instagram para el entendimiento del comportamiento de los usuarios?

PI3: Qué técnicas de preprocesamiento se utilizan comúnmente en los estudios sobre la predicción de la popularidad en Instagram?

PI4: ¿Cuáles son las brechas en la literatura sobre la predicción de la popularidad en Instagram?

Además, el proceso de identificación de los artículos se lo elaboró en cuatro pasos, como se refleja en la Figura 1, los cuales son explicados a continuación.

Figura1. Flujograma de PRISMA relacionado con los documentos académicos relacionados con modelos predictivos en Instagram.



El primer paso se refiere a una identificación inicial, donde se

procedió a realizar una búsqueda en bases de datos académicas,

incluyendo a Scopus, Web of Science, Taylor & Francis, IEEE y Springer. En esta búsqueda, se seleccionaron estudios que guardaran relación con los modelos predictivos y las tendencias en Instagram y otras redes sociales, usando cadena de búsqueda con palabras claves incluyendo como modelo predictivo, predicción, Instagram, popularidad, aprendizaje automático. Estas palabras también fueron usadas en inglés y se restringió la búsqueda a artículos publicados dentro de los cinco últimos años, lo que reflejó un total de 1008 documentos. El segundo paso involucró la selección de documentos, donde se eliminaron aquellos que no eran relevantes al caso de estudio revisando los títulos y resúmenes. Por ejemplo, los que incluían la palabra Instagram, pero para un contexto general que no eran de consideración para el alcance de esta revisión. O también aquellos documentos que no fueran artículos científicos de revistas y conferencias, libros o capítulos de libros o que no tuvieran metodología cuantitativa. Con base en esta selección quedaron 596 documentos.

Posteriormente, durante la etapa de legibilidad se procedió a elegir los documentos académicos que aportaban mayor información al tema de tendencias de Instagram y otras redes sociales con modelos predictivos, analizándolos en mayor detalle. En esta etapa se aplicó un análisis riguroso para incluir y excluir estudios que no cumplieran con los criterios metodológicos establecidos en este estudio. Después de este proceso, se seleccionaron 103 documentos académicos. Finalmente, durante la fase de inclusión se tomaron en consideración los estudios que cumplieron con los criterios de inclusión por relacionarse directamente con tendencias de popularidad en Instagram. Algunos estudios fueron excluidos debido a que, a pesar de ser recientes, se centran en aspectos que han sido superados. De este modo, este estudio contempló finalmente 43 estudios sobre la popularidad de Instagram y otras redes sociales usando modelos predictivos y métodos cuantitativos. Los criterios de inclusión y exclusión usados en este estudio se ven reflejados en la Tabla 1 a continuación.

Tabla 1. Criterios de inclusión y exclusión implementados en este estudio. Fuente: Las autoras.

Criterio	Inclusión	Exclusión
Tipo de estudios	Artículos revisados por pares académicos, libros, capítulos de libros y publicaciones en conferencias sobre machine learning, análisis de datos, tecnología, publicados en inglés y en español en Scopus, Web of Science, Taylor & Francis, IEEE y Springer.	Estudios que no hayan pasado por revisión por pares, como blogs, informes, o material de fuente que no sean de ámbito académico. Se excluyeron tesis de pregrado y posgrado.
Fecha de publicación	Estudios publicados en los últimos 5 años, que abarca desde el 2019 hasta el 2024.	Estudios realizados a años anteriores al 2019.
Relevancia del tema	Artículos que aborden la popularidad de Instagram y otras redes sociales con modelos predictivos o métodos cuantitativos, como algoritmos, métricas, análisis de datos.	Artículos que no se centren en la popularidad de Instagram u otras redes sociales y que se enfocan solo en metodología cualitativa.
Metodología	Estudios que definan claramente la metodología, y con datos que respalden sus conclusiones.	Estudios que en la metodología contiene deficiencia o falta de información.

Para el análisis se usó la técnica de análisis de contenido mixto para examinar las principales tendencias y patrones en la literatura sobre la predicción de popularidad en Instagram. Este enfoque es relevante en el contexto de las redes sociales, donde la medición de variables cuantitativas de interacción, como me gusta, comentarios, clics y menciones, y las cualitativas como el alcance y la influencia requiere una comprensión integral de los comportamientos y motivaciones de los usuarios. El

análisis de contenido mixto combina dos enfoques para estudiar cómo las personas se comunican en redes sociales. Primero, usa métodos cualitativos para entender en profundidad el significado detrás de los mensajes y la interacción. Luego, incorpora métodos cuantitativos para medir y contar las diferentes formas de comunicación. Juntas, estas dos perspectivas proporcionan una imagen más completa de los patrones y tendencias en las redes sociales (Grassia et al., 2024).

3. Resultados y discusión

Esta sección presenta los principales hallazgos de la revisión de la literatura sobre la predicción de la popularidad de Instagram, de acuerdo con las PI planteadas anteriormente. Además, en la Tabla

2, se resumen las investigaciones más relevantes usados en este estudio, destacando sus contribuciones, enfoques metodológicos y hallazgos sobre la eficacia de modelos predictivos en Instagram.

Tabla 2. Resumen de estudios usados en esta investigación de modelos predictivos. Fuente: Las autoras.

Autores	Modelo predictivo	Técnicas de preprocesamiento	Red social	Variables predictoras	Tipo de publicación usada en el estudio	Contribución
Adiningtyas & Auliani (2024)	Latent Dirichlet Allocation (LDA).	• Conversión a minúsculas • Eliminación de caracteres no alfabéticos • Tokenización • Eliminación de stopwords y stemming.	Instagram	Valor emocional, social y calidad.	Comentarios	El modelo LDA identificó nueve temas principales que los usuarios mencionaban con frecuencia en sus comentarios. Estos temas abarcan problemas con el servicio, como fallos en las transferencias bancarias, en la actualización de cuentas, en las promociones de cashback, y también aspectos positivos, como la recepción de incentivos del programa gubernamental Prakerja.
Aldous et al. (2023)	AdaBoost, Decision Tree y Random Forest.	• Term Frequency-Inverse Document Frequency (TF-IDF) • Tokenización y lematización • Reducción de ruido.	• Facebook • Instagram • Twitter • YouTube • Reddit	Características del lenguaje, derivadas de TF-IDF. Características de metadatos, incluidos vectores de temas, longitud del contenido, emojis, puntuación y análisis de sentimiento.	Publicaciones de noticias de 53 medios de comunicación	El estudio desarrolló un marco de interacción de cuatro niveles para clasificar la interacción del usuario desde la interacción privada hasta el intercambio público. También proporcionó modelos predictivos para ayudar a los creadores de contenido a optimizar la interacción en función del análisis de sentimientos.
Chakrabarti et al. (2023)	Support Vector Regression, Extreme Gradient Boosting, Random Forest y aprendizaje profundo personalizado.	• Eliminación de publicaciones no relacionadas, en idiomas distintos al inglés • Técnicas de extracción de hashtags y palabras clave utilizando modelos como KeyBERT.	Twitter	Métricas de interacción, como me gusta, comentarios, compartidos, el número de seguidores del usuario, y los hashtags utilizados en las publicaciones.	Texto de tweets	El método propuesto superó a las técnicas existentes al mejorar la precisión de las recomendaciones de hashtags, logrando una tasa de éxito del 82.81% en publicaciones de baja popularidad y 18.81% en las de popularidad media.
Chaudhary et al. (2021)	Arboles de decisión supervisados, Regresión Lineal, Bosques Aleatorios,	• Limpieza de datos eliminando valores perdidos y registros	• Facebook • Twitter • LinkedIn • YouTube • Instagram • Pinterest	Me gusta, seguidores, visitas y descargas de los productos en cada plataforma social.	Fotos y videos	El modelo predijo el comportamiento de los consumidores con una precisión del 98% en datos de

	Regresión con AdaBoost y XGBoost	- duplicados - Técnicas de detección de datos atípicos y se eliminaron datos ruidosos.				validación, siendo el árbol de decisión el más preciso.
Chiu et al. (2021)	Redes Neuronales Convolucionales (CNN), Word2Vec, Random Forest	- Decodificación de emoticones - Eliminación de elementos redundantes - Segmentación de palabras.	Instagram	- Imágenes: Características visuales extraídas con CNN - Texto: Características semánticas extraídas con Word2Vec y LSTM - Comportamiento: Número de me gusta, tiempo entre publicaciones, localización, hora y día de la publicación.	Imágenes y textos	El estudio logró una medida F1 de 0.835 en la detección de usuarios depresivos. La combinación multimodal y el enfoque de agregación temporal demostraron mejorar la precisión en la detección de tendencias depresivas.
El Bahy et al. (2024)	Big Five	No se especifica.	Instagram	- Visual: Color y formas - Emocional: Sentimiento asociado a la imagen - Contenido: Tipo de publicación.	Fotos	Los resultados muestran que las predicciones de personalidad obtuvieron buenos resultados, con mayor precisión en la predicción de rasgos como conciencia para las mujeres y apertura para los hombres. También se evaluó el impacto de la cantidad de imágenes en la precisión de la predicción
Khairy et al. (2024)	Multinomial Naive Bayes, Bernoulli Naive Bayes, Regresión Logística, Clasificador de Gradiente Estocástico (SGD), SVC, Linear SVC, Árbol de Decisión y Random Forest.	- Limpieza de datos eliminando palabras vacías - Normalización, stemming, y transformación de texto en datos numéricos mediante TF-IDF.	- Twitter - YouTube - Kaggle	Características extraídas del texto, principalmente mediante TF-IDF.	Comentarios y publicaciones de texto	Se encontró que el rendimiento de las técnicas de remuestreo mejora el rendimiento de los clasificadores en datos desequilibrados. Entre las técnicas más efectivas están SMOTE y SMOTE combinado con Tomek Links. Además, el clasificador Random Forest logró un desempeño superior con la técnica de random oversampling.
Khorrami et al. (2022)	Árboles extra, árboles gradientes y bosques aleatorios.	No se especifica.	Instagram	Color de las fotos.	Imágenes	Los resultados mostraron que cada rasgo de personalidad tiene un tipo especial de preferencia de color y publicación de fotografías.
Kumar & Vohra (2023)	Algoritmos de aprendizaje profundo.	- Eliminación de datos duplicados y manejos de datos nulos - Normalización y escalas - Tokenización y estandarización	Instagram	Características visuales, emocionales y de contenido.	Imágenes	El estudio destaca que aprovechar el aprendizaje profundo y el análisis de metadatos es crucial para avanzar en la sostenibilidad tecnológica en el sector turístico.
Lekkas et al. (2021)	Modelo de aprendizaje automático en ensamblaje.	Estandarización de datos.	Instagram	- Actividad: Número de seguidores, número de seguidos, cantidad de publicaciones, promedio de comentarios y "me gusta" por publicación - Lingüísticas: Cantidad de	Imágenes	El modelo ensamblado logró una precisión del 70.2% en la predicción de la ideación suicida aguda. Las variables de actividad en Instagram, como el número de seguidores y la interacción, tuvieron mayor importancia predictiva que las características lingüísticas. El uso de palabras asociadas a

				palabras, uso de pronombres en primera persona, palabras de afecto, emociones negativas, procesos cognitivos y facilidad de lectura Derivadas: Relación seguidores/seguídos, nivel de interacción (engagement), promedio de comentarios y "me gusta" por seguidor.		emociones negativas también fue relevante en la predicción
Saeidi & Baradari (2023)	SVM, CART (Árboles de Clasificación y Regresión), y Random Forest	- Limpieza y transformación de los datos - Asignación de número a variables categóricas - Eliminación de datos.	Instagram	Número de seguidores, número de likes por publicación, número de comentarios por publicación, y número de publicaciones	Publicaciones en general	El estudio encontró que existe una correlación inversa entre el número de publicaciones y el número de me gusta. También mostró que el número de seguidores no impacta en la cantidad de me gusta ni en los comentarios. El modelo propuesto alcanzó una precisión del 94% en la detección de páginas falsas.
Yu & Egger (2021)	SVM y Bosque aleatorio	Etiquetas y Agrupación de imágenes.	Instagram	Composiciones de colores	Foto única	Las combinaciones de colores afectan el engagement en Instagram: tonos azules aumentan el engagement en fotos de naturaleza, mientras que el rojo lo hace en imágenes de gastronomía
Zhong et al. (2020)	Neural Link Prediction (NeuLP)	Codificación de redes neuronales de grafos para agregar información local de las conexiones del usuario y actualizar sus representaciones embebidas.	Instagram	Atributos de ubicación (check-ins) y hashtags. Variables estructurales de la red, Interacciones lineales y no lineales.	Usuarios que realizaron al menos 100 check-ins de ubicación y tenían un mínimo de 10 amigos.	NeuLP demostró ser efectivo en la predicción de enlaces sociales en comparación con otros modelos basados en redes neuronales de grafos, logrando hasta un 5.8% de mejora en la métrica AUC. Además, el modelo es flexible y permite la predicción de enlaces utilizando múltiples tipos de atributos de usuarios.

PI1: ¿Cómo utilizan las organizaciones los modelos predictivos basados en datos de Instagram?

Instagram, como una de las plataformas de redes sociales más populares a nivel mundial, ha experimentado un crecimiento exponencial en el número de usuarios y en la cantidad de datos consumidos y generados. Los

usuarios de esta red social están expuestos a publicaciones relacionados con fotos individuales, carruseles y reels. Este crecimiento ha creado la necesidad de desarrollar herramientas y técnicas avanzadas para analizar y aprovechar estos datos. Debido a esta necesidad, los modelos predictivos han emergido como una herramienta clave para extraer valor

de la cantidad de datos y tener ventajas competitivas en el ámbito organizacional (Gon, 2021). Los modelos predictivos varían en complejidad y metodología. Los más utilizados son los modelos de redes neuronales, SVM, árboles de decisión, y los métodos basados en regresión. Estos modelos han demostrado ser eficaces para predecir la popularidad de una publicación utilizando diversas métricas, como me gusta, comentarios, acciones y seguidores.

Un aspecto relevante para la popularidad de la adopción de Instagram se basa en que puede guiar la toma de decisiones basadas en datos. Esta plataforma se usa para la personalización de contenido con base en las necesidades del público objetivo y para la predicción de tendencias emergentes. Por ejemplo, se ha demostrado su efectividad en el entendimiento de las necesidades de los usuarios de sectores demográficos específicos (Gupta et al., 2021; Hazari & Sethna, 2023), en la personalización de experiencias para efectos de marketing e intención de compra (Cavdar Aksoy et al., 2021; Hussain et al., 2023). Dentro del campo

médico, Instagram ha sido ampliamente usado para estudios psicológicos (Chaudhary et al., 2021; El Bahy et al., 2024; Lekkas et al., 2021), mientras que para el turismo se ha usado para la predicción de sitios de interés para turistas (Sohrabi et al., 2020; Son & Park, 2023). También estudios previos se han enfocado en predicciones de sostenibilidad (Kumar & Vohra, 2023; Mahmoud et al., 2022) que permiten tener un entendimiento más integral sobre la aplicabilidad de modelos predictivos usando los datos de Instagram, lo cual refuerza su versatilidad como una herramienta predictiva y analítica en diversos campos.

Sin embargo, es necesario puntualizar las limitaciones que pueden surgir del uso de los datos de Instagram para efectos predictivos. Quizá uno de los más importantes se refiere a la calidad y veracidad de los datos debido a su influencia en las predicciones (Sallah et al., 2023; Zachlod et al., 2022). Es decir, si los datos están incompletos, incorrectos o con ruido, el modelo podría producir predicciones erróneas o insuficientes. Además, el uso de datos de plataformas de redes

sociales conlleva preocupaciones éticas, especialmente en relación con la privacidad de los usuarios (Chaudhary et al., 2021; Fernández-Rovira, 2022). Ante esta situación es relevante garantizar que se cumplen con las regulaciones de privacidad y se tiene el consentimiento adecuado para el manejo de datos de otros usuarios. Bajo este principio, actualmente Instagram no permite realizar procesos de minería de datos ni raspado de web debido a consideraciones éticas y legales. Otra limitación es la transparencia con la cual los algoritmos desarrollan las predicciones (Williams et al., 2023), así como la interpretabilidad. Por ejemplo, se consideran a algoritmos como las redes neuronales o máquinas de vectores de soporte como cajas negras puesto que sus procesos internos son difíciles de interpretar o comprender. Esto ha evidenciado la necesidad de la transparencia y explicabilidad de los modelos predictivos usando algoritmos como razonamiento evidencial (Fang et al., 2021) que permite a las personas entender cómo los algoritmos toman decisiones. Finalmente, los estudios que miden la popularidad suelen centrarse en métricas como me

gusta y comentarios, las cuales pueden tener una capacidad limitada para reflejar la calidad real del contenido. Además, estas métricas pueden ser manipuladas mediante prácticas engañosas, como el uso de cuentas automatizadas o la compra de seguidores. Por esta razón, es importante realizar análisis cualitativos que permita proporcionar un análisis más profundo de la interacción de los usuarios.

PI2: ¿Qué tipos de publicaciones y métricas son más efectivas en los modelos predictivos de popularidad en Instagram para el entendimiento del comportamiento de los usuarios?

Como se mencionó anteriormente, Instagram se compone tradicionalmente de publicaciones relacionadas con fotos individuales, carruseles y reels. Adicionalmente, también están las historias que son publicaciones temporales que desaparecen después de 24 horas y videos de una duración mayor a 60 minutos, que anteriormente se conocía como IGTV. Uno de los tipos de publicaciones que más se ha usado para modelaciones son las fotos individuales. Por ejemplo, para predecir suicidios en adolescentes

(Lekkas et al., 2021), personalidad de usuarios extrayendo aspectos visuales, emocionales y de contenido (El Bahy et al., 2024; Khorrami et al., 2022), o también para predicción de la popularidad de marcas de alta moda (Javadian Sabet et al., 2021). Sin embargo, no existen mayores investigaciones relacionadas con el uso de carruseles o reels debido a que tienden a ser más complejos para análisis, en comparación con las fotos individuales. Dado que un carrusel puede contener hasta diez fotos o videos, la predicción se vuelve más compleja, ya que no es fácil identificar qué elemento está influyendo en las predicciones. De igual manera, el análisis de reels requiere enfoques analíticos variados para lograr un análisis eficiente e integral.

Por otro lado, las métricas usadas en los modelos predictivos varían dependiendo de la naturaleza de las investigaciones, sobre todo cuando se refiere a la popularidad de las publicaciones. Este concepto se refiere al grado de interacción y visibilidad recibidas (Yang et al., 2020). Tradicionalmente, este concepto se ha asociado a métricas

como el número de me gusta que representa la aprobación, los comentarios que reflejan la interacción de los usuarios, los compartidos que mide la difusión de la publicación. También las impresiones el cual es un indicador que mide el número de veces que la publicación es vista, la tasa de clics que simboliza las acciones de los usuarios frente a enlaces en las publicaciones, el número de seguidores y las reacciones obtenidas en las historias. Se ha evidenciado, además, que el uso de hashtags, emojis y menciones, por citar algunos, influye en la popularidad de las publicaciones (Chakrabarti et al., 2023; Siever et al., 2020). Sin embargo, también hay críticas al uso exclusiva de este tipo de métricas debido a que no necesariamente reflejan una publicación de calidad, por interacciones superficiales, sesgo de popularidad o incluso debido a la manipulación de ellas mediante el uso de cuentas automáticas (Cuevas-Molano et al., 2021; Shin & Lee, 2020; Tafesse & Wood, 2021).

PI3: ¿Qué técnicas de preprocesamiento se utilizan comúnmente en los estudios

sobre la predicción de la popularidad en Instagram?

En la construcción de modelos predictivos, las técnicas de preprocesamiento son fundamentales para optimizar el rendimiento de los algoritmos, ya que maximizan la calidad del proceso, así como la mejora en la precisión y eficacia de las predicciones. Entre estas técnicas se encuentra la limpieza de los datos, que implica la reducción de ruido, dimensionalidad, valores atípicos y datos incompletos (Khairy et al., 2024; Khan & Ali, 2024). En el contexto de Instagram esto puede significar la eliminación de registros con datos faltantes, imputación de datos, corrección de inconsistencias en el etiquetado, remoción o transformación de datos que no aportan valor para los procesos de predicción (Adiningtyas & Auliani, 2024; Gusain et al., 2025; Rahman et al., 2024).

Otra técnica que se ha usado es la transformación de características, conocidas en inglés como one-hot encoding. Incluye la conversión de datos categóricos en formatos numéricos mediante la codificación de etiquetas (Zhong et al., 2020). En

Instagram se puede aplicar a variables relacionadas con el tipo de contenido que se publica o con las etiquetas de usuario, lo cual permite que el modelo predictivo lo interprete de forma más adecuada. Además, la normalización y estandarización de datos numéricos, como la cantidad de seguidores o frecuencia de publicaciones, es importante para evitar que las características con escalas más grandes dominen al modelo (Aldous et al., 2023). Por otro lado, técnicas relacionadas con el manejo de datos textuales también son relevantes en el contexto de los datos de esta plataforma, donde las descripciones y comentarios son una fuente valiosa de información. Por ejemplo, la tokenización, lematización y eliminación de palabras vacías son usadas para transformar texto crudo en un formato que pueda ser analizado por algoritmos de aprendizaje automático. Y, el análisis de sentimiento y extracción de palabras claves a través de modelos de lenguaje natural son esenciales para capturar el contenido semántico del texto en las publicaciones.

PI4: ¿Cuáles son las brechas en la literatura sobre la predicción de la popularidad en Instagram?

Se identificaron varias brechas que requieren atención dentro del área de modelos predictivos. Primero, es crucial desarrollar modelos predictivos utilizando algoritmos transparentes y explicables. La falta de transparencia en estos modelos puede generar desconfianza entre los usuarios y perpetuar sesgos, afectando la equidad en el alcance y visibilidad de las publicaciones. Para abordar esta brecha, futuras investigaciones deberían centrarse en crear herramientas que no solo sean efectivas en sus predicciones, sino que también ofrezcan visualizaciones claras y comprensibles de cómo se generan esos resultados. Esto podría incluir la implementación de técnicas de explicabilidad, como LIME (Local Interpretable Model-agnostic Explanations) o SHAP (SHapley Additive exPlanations), que ayudan a los usuarios a entender qué características específicas influyen en la popularidad de sus publicaciones.

Segundo, es fundamental comprender las implicaciones éticas

y de privacidad, especialmente dado el uso creciente de datos sensibles. A medida que las plataformas sociales recopilan más información personal, surge la necesidad de establecer prácticas éticas que protejan la privacidad de los usuarios. Las investigaciones futuras deberían centrarse en desarrollar directrices claras para la recolección y uso de datos, creando marcos que integren principios éticos en el diseño de modelos predictivos. Esto podría incluir la colaboración con expertos en ética y derechos digitales para asegurar que los modelos cumplan con normativas legales y respeten la dignidad de los usuarios.

Tercero, aunque existen estudios que se han enfocado en la predicción de la popularidad a través de fotografías individuales, es esencial expandir este enfoque para incluir formatos emergentes como carruseles y reels. Estos tipos de contenido son cada vez más populares y su dinámica puede influir en la interacción de los usuarios. La investigación futura debería explorar cómo estos formatos afectan las métricas de popularidad y qué elementos específicos, como la

secuencia de imágenes o la duración de los videos, son más efectivos. Implementar estudios de caso y análisis comparativos podría proporcionar una comprensión más profunda de los factores que impulsan la popularidad en estos nuevos formatos. Además, la interacción entre usuarios y algoritmos representa un área poco explorada, y comprender cómo los usuarios ajustan sus estrategias de contenido en respuesta a cambios en las políticas de Instagram puede proporcionar resultantes relevantes para las organizaciones. También es crucial considerar el impacto de las tendencias sociales y culturales en la popularidad de las publicaciones, investigando cómo movimientos sociales y eventos culturales influyen en la participación del público.

4. Conclusiones

Este estudio se ha enfocado en realizar una revisión de la literatura sobre la predicción de la popularidad de Instagram examinando métodos y técnicas para predecir el éxito de publicaciones. Los resultados evidencian que Instagram es una fuente relevante de datos para el desarrollo de modelos predictivos y,

cuando son usados de forma adecuada, pueden dar ventajas competitivas a las organizaciones. También se identificó como la literatura ha usado los tipos de publicaciones, métricas para evaluar el rendimiento de los algoritmos y las técnicas de preprocesamiento. Finalmente, se evidenciaron brechas relacionadas con la transparencia y explicabilidad de los algoritmos, consideraciones éticas y de privacidad de los datos y el uso de carruseles y reels para efectos predictivos.

Con relación a la limitación del estudio podemos identificar dos. La primera se relaciona con los idiomas seleccionados para las búsquedas de los documentos académicos que se centraron en inglés y español, mientras que la segunda limitación se relaciona con las bases de datos académicas seleccionadas, debido a que podrían existir estudios relevantes publicados en otras bases de datos. Finalmente, futuros estudios pueden centrarse en el desarrollo de modelos predictivos transparentes y explicables, los cuales son cada vez más necesarios para comprender cómo los algoritmos toman decisiones. Es

fundamental explorar las implicaciones éticas y de privacidad que pueden surgir en el proceso de obtención de datos y modelado. Además, se debe investigar la influencia de diferentes tipos de publicaciones, como carruseles y reels, en los modelos predictivos, así como el impacto de movimientos sociales y culturales en la participación del público y en la popularidad de las publicaciones en Instagram.

Bibliografía

- Adiningtyas, H., & Auliani, A. S. (2024). Customer perceived value: A study based on customer perception on social media. *Seventh Information Systems International Conference (ISICO 2023)*, 234, 1451–1458. <https://doi.org/10.1016/j.procs.2024.03.145>
- Aldous, K. K., An, J., & Jansen, B. J. (2023). What really matters? Characterising and predicting user engagement of news postings using multiple platforms, sentiments and topics. *Behaviour & Information Technology*, 42(5), 545–568. <https://doi.org/10.1080/0144929X.2022.2030798>
- Casaló, L. V., Flavián, C., & Ibáñez-Sánchez, S. (2021). Be creative, my friend! Engaging users on Instagram by promoting positive emotions. *Journal of Business Research*, 130, 416–425. <https://doi.org/10.1016/j.jbusres.2020.02.014>
- Cavdar Aksoy, N., Tumer Kabadayi, E., Yilmaz, C., & Kocak Alan, A. (2021). A typology of personalisation practices in marketing in the digital age. *Journal of Marketing Management*, 37(11–12), 1091–1122. <https://doi.org/10.1080/0267257X.2020.1866647>
- Chakrabarti, P., Malvi, E., Bansal, S., & Kumar, N. (2023). Hashtag recommendation for enhancing the popularity of social media posts. *Social Network Analysis and Mining*, 13(1), 21. <https://doi.org/10.1007/s13278-023-01024-9>
- Chaudhary, K., Alam, M., Al-Rakhmi, M. S., & Gumaei, A. (2021). Machine learning-based mathematical modelling for prediction of social media consumer behavior using big data analytics. *Journal of Big Data*, 8(1), 73. <https://doi.org/10.1186/s40537-021-00466-2>
- Chiu, C. Y., Lane, H. Y., Koh, J. L., & Chen, A. L. P. (2021).

- Multimodal depression detection on instagram considering time interval of posts. *Journal of Intelligent Information Systems*, 56(1), 25–47.
<https://doi.org/10.1007/s10844-020-00599-5>
- Cuevas-Molano, E., Matosas-López, L., & Bernal-Bravo, C. (2021). Factors increasing consumer engagement of branded content in Instagram. *IEEE Access*, 9, 143531–143548.
<https://doi.org/10.1109/ACCESS.2021.3121186>
- Diefenbach, S., & Anders, L. (2022). The psychology of likes: Relevance of feedback on Instagram and relationship to self-esteem and social status. *Psychology of Popular Media*, 11(2), 196–207.
<https://doi.org/10.1037/ppm0000360>
- Dixon, S. J. (2024). Instagram: Number of global users 2020-2025. Statista.
<https://www.statista.com/statistics/183585/instagram-number-of-global-users/>
- El Bahy, S., Aboutabit, N., & Hafidi, I. (2024). Analysis and prediction of personality traits using a self-generated database of Moroccan instagram users: Impact of gender on image content and quantity on prediction accuracy. *Multimedia Tools and Applications*.
<https://doi.org/10.1007/s11042-024-19101-2>
- Fang, R., Liao, H., Yang, J.-B., & Xu, D.-L. (2021). Generalised probabilistic linguistic evidential reasoning approach for multi-criteria decision-making under uncertainty. *Journal of the Operational Research Society*, 72(1), 130–144.
<https://doi.org/10.1080/01605682.2019.1654415>
- Fernández-Rovira, C. (2022). Ethical insights for the social media age. In *Predictive Technology in Social Media* (First, pp. 158–179). CRC Press.
- Gon, M. (2021). Local experiences on Instagram: Social media data as source of evidence for experience design. *Journal of Destination Marketing & Management*, 19, 100435.
<https://doi.org/10.1016/j.jdmm.2020.100435>
- Grassia, V., d'Apuzzo, F., Alansari, R. A., Jamilian, A., Sayahpour, B., Adel, S. M., & Nucci, L. (2024). Instagram and clear aligner therapy: A content analysis of patient perspectives. *Seminars in Orthodontics*.
<https://doi.org/10.1053/j.sodo.2024.05.009>
- Gupta, M., Sinha, N., Singh, P., & Liébana-Cabanillas, F. (2021). Instagram advertising

- among young consumers in wearable fitness trackers: The moderating role of technology acceptance factors. *Journal of Global Marketing*, 34(5), 411–432.
<https://doi.org/10.1080/08911762.2021.1931616>
- Gusain, R., Shekhar, S., Vidyarthi, A., Prakash, R., & Gowri, R. (2025). Smart healthcare system using machine learning and IoT: Vaccine delivery and monitoring. En *Embedded Devices and Internet of Things* (pp. 153–177). CRC Press.
- Hazari, S., & Sethna, B. N. (2023). A comparison of lifestyle marketing and brand influencer advertising for generation Z Instagram users. *Journal of Promotion Management*, 29(4), 491–534.
<https://doi.org/10.1080/10496491.2022.2163033>
- Hussain, A., Hooi Ting, D., Zaib Abbasi, A., & Rehman, U. (2023). Integrating the S-O-R model to examine purchase intention based on Instagram sponsored advertising. *Journal of Promotion Management*, 29(1), 77–105.
<https://doi.org/10.1080/10496491.2022.2108185>
- Ibrahim, B., & Aljarah, A. (2023). The era of Instagram expansion: Matching social media marketing activities and brand loyalty through customer relationship quality. *Journal of Marketing Communications*, 29(1), 1–25.
<https://doi.org/10.1080/13527266.2021.1984279>
- Javadian Sabet, A., Brambilla, M., & Hosseini, M. (2021). A multi-perspective approach for analyzing long-running live events on social media. A case study on the “Big Four” international fashion weeks. *Online Social Networks and Media*, 24, 100140.
<https://doi.org/10.1016/j.osnem.2021.100140>
- Khairy, M., Mahmoud, T. M., & Abd-El-Hafeez, T. (2024). The effect of rebalancing techniques on the classification performance in cyberbullying datasets. *Neural Computing and Applications*, 36(3), 1049–1065.
<https://doi.org/10.1007/s00521-023-09084-w>
- Khan, A., & Ali, R. (2024). Unraveling minds in the digital era: A review on mapping mental health disorders through machine learning techniques using online social media. *Social Network Analysis and Mining*, 14(1), 78.
<https://doi.org/10.1007/s13278-024-01205-0>
- Khorrami, M., Khorrami, M., & Farhangi, F. (2022). Evaluation of tree-based ensemble algorithms for

- predicting the big five personality traits based on social media photos: Evidence from an Iranian sample. *Personality and Individual Differences*, 188, 111479.
<https://doi.org/10.1016/j.paid.2021.111479>
- Kumar, A., & Vohra, R. (2023). Impact of deep learning models for technology sustainability in tourism using big data analytics. En V. Kadyan, T. P. Singh, & C. Ugwu (Eds.), *Deep Learning Technologies for the Sustainable Development Goals: Issues and Solutions in the Post-COVID Era* (pp. 83–96). Springer Nature Singapore.
https://doi.org/10.1007/978-981-19-5723-9_6
- Lekkas, D., Klein, R. J., & Jacobson, N. C. (2021). Predicting acute suicidal ideation on Instagram using ensemble machine learning models. *Internet Interventions*, 25, 100424.
<https://doi.org/10.1016/j.invent.2021.100424>
- Mahmoud, A. B., Ball, J., Rubin, D., Fuxman, L., Mohr, I., Hack-Polay, D., Grigoriou, N., & Wakibi, A. (2022). Pandemic pains to Instagram gains! COVID-19 perceptions effects on behaviours towards fashion brands on Instagram in Sub-Saharan Africa: Tech-native vs non-native generations. *Journal of Marketing Communications*, 28(8), 864–888.
<https://doi.org/10.1080/13527266.2021.1971282>
- Mohammed, A., & Kora, R. (2023). A comprehensive review on ensemble deep learning: Opportunities and challenges. *Journal of King Saud University - Computer and Information Sciences*, 35(2), 757–774.
<https://doi.org/10.1016/j.jksuci.2023.01.014>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Alonso-Fernández, S. (2021). Declaración PRISMA 2020: Una guía actualizada para la publicación de revisiones sistemáticas. *Revista Española de Cardiología*, 74(9), 790–799.
<https://doi.org/10.1016/j.recesp.2021.06.016>
- Rahman, T., Shahrin, R., Pospu, F. A., Sultana, N., & Rahman, R. M. (2024). Text-based data analysis for mental health using explainable AI and deep learning. En R. Lee (Ed.),

- Networking and Parallel/Distributed Computing Systems: Volume 18 (pp. 17–31). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-53274-0_3
- Saeidi, S., & Baradari, Z. (2023). Examining the correlation between metrics in the Instagram social network to identify fake pages and improve marketing. *Computers in Human Behavior Reports*, 12, 100341. <https://doi.org/10.1016/j.chbr.2023.100341>
- Sallah, A., Alaoui, E. A. A., & Agoujil, S. (2023). Interpretability based approach to detect fake profiles in Instagram. En M. Ben Ahmed, B. A. Abdelhakim, B. K. Ane, & D. Rosiyadi (Eds.), *Emerging Trends in Intelligent Systems & Network Security* (pp. 306–314). Springer International Publishing.
- Shin, J., & Lee, S. (2020). Intimacy between actual users and virtual agents: Interaction through “likes” and “comments”. 2020 14th International Conference on Ubiquitous Information Management and Communication (IMCOM), 1–4. <https://doi.org/10.1109/IMCO M48794.2020.9001810>
- Siever, C. M., Siever, T., & Stöckl, H. (2020). Emoji-text relations on Instagram: Empirical corpus studies on multimodal uses of the iconographic mode. En *Shifts toward Image-Centricity in Contemporary Multimodal Practices* (pp. 177–203). Routledge.
- Sohrabi, B., Raeesi Vanani, I., Nasiri, N., & Ghassemi Rudd, A. (2020). A predictive model of tourist destinations based on tourists’ comments and interests using text analytics. *Tourism Management Perspectives*, 35, 100710. <https://doi.org/10.1016/j.tmp.2020.100710>
- Son, H., & Park, Y. E. (2023). Predicting user engagement with textual, visual, and social media features for online travel agencies’ Instagram post: Evidence from machine learning. *Current Issues in Tourism*, 1–15. <https://doi.org/10.1080/13683500.2023.2278087>
- Tafesse, W., & Wood, B. P. (2021). Followers’ engagement with instagram influencers: The role of influencers’ content and engagement strategy. *Journal of Retailing and Consumer Services*, 58, 102303. <https://doi.org/10.1016/j.jretconser.2020.102303>
- Williams, D., Khan, E. M., Pathak, N., & Srivastava, J. (2023). Social

- p value: A computational model for measuring influence on purchases and actions for individuals and systems.
- Journal of Advertising*
- , 52(2), 247–263.
-
- <https://doi.org/10.1080/00913367.2021.2002743>
- Wu, C.-C., Hsu, P.-Y., Xu, N., Cheng, M.-S., & Chen, Y.-Y. (2022). Prediction of personality traits through Instagram photo HSV. En M. Kurosu (Ed.), *Human-Computer Interaction. Technological Innovation* (pp. 279–287). Springer International Publishing.
- Yang, K., Kim, H. M., & Tanoff, L. (2020). Signaling trust: Cues from Instagram posts. *Electronic Commerce Research and Applications*, 43, 100998.
<https://doi.org/10.1016/j.elera.2020.100998>
- Yu, J., & Egger, R. (2021). Color and engagement in touristic Instagram pictures: A machine learning approach. *Annals of Tourism Research*, 89, 103204.
<https://doi.org/10.1016/j.annals.2021.103204>
- Zachlod, C., Samuel, O., Ochsner, A., & Werthmüller, S. (2022). Analytics of social media data: State of characteristics and application. *Journal of Business Research*, 144, 1064–1076.
<https://doi.org/10.1016/j.jbusres.2022.02.016>
- Zhong, Z., Zhang, Y., & Pang, J. (2020). NeuLP: An end-to-end deep-learning model for link prediction. En Z. Huang, W. Beek, H. Wang, R. Zhou, & Y. Zhang (Eds.), *Web Information Systems Engineering – WISE 2020* (pp. 96–108). Springer International Publishing.