

Base de Datos

Juan Felipe Salamanca González

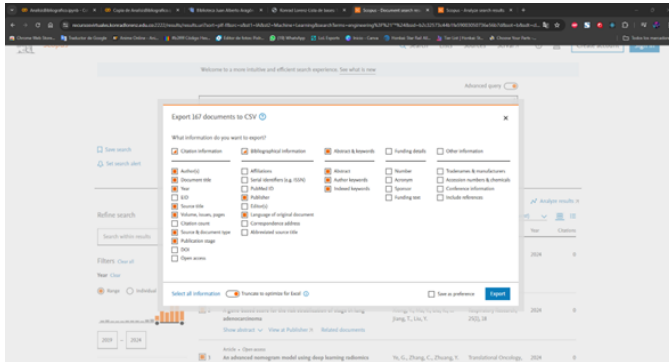


Fig. 1. Creacion archivo .CSV

Abstract—El informe mostrara los pasos que utilizamos y el analisis bibliografico a mas de 100 articulos sobre inteligencia artificial y el aprendizaje automatico por lo que para nuestro analisis haremos uso de Google Colab y librerias de python para hacer un analisis grafico y poder establecer las conclusiones que se veran mas adelante

Keywords: Plotly, Python, Google Colab, IA, aprendizaje automatico, Scopus, Pandas

I. INTRODUCTION

El objetivo fue realizar un análisis bibliográfico exhaustivo sobre inteligencia artificial (IA) y aprendizaje automático (machine learning), utilizando una base de datos de la Universidad y herramientas como Google Colab y un archivo CSV que generamos apoyandonos de la base de datos Scopus y cargaremos en Google Colab con pandas. Se analizaron 106 artículos de la Fundación Universitaria Konrad Lorenz. Este análisis permitió comprender rápidamente los títulos, palabras clave, resúmenes y fuentes de los artículos, facilitando la identificación de información relevante en el área de interés.

II. METODOLOGÍA

La metodología aplicada para el análisis y visualización del texto se desarrolló utilizando una serie de bibliotecas de Python. En primer lugar, se utilizó 'pandas' para la manipulación de datos, permitiendo la carga y exploración inicial del conjunto de datos proveniente del archivo 'scopus.csv'. Posteriormente, se empleó 'nltk' (Natural Language Toolkit) para realizar la tokenización de oraciones y palabras, así como para eliminar las palabras vacías utilizando la lista de stopwords en inglés descargada a través de la función 'nltk.download('stopwords')'. Este paso de preprocesamiento aseguró que el análisis se centrara en términos relevantes.

En la siguiente etapa, se generarán visualizaciones para comprender mejor los datos. Se utilizaron gráficos de barras

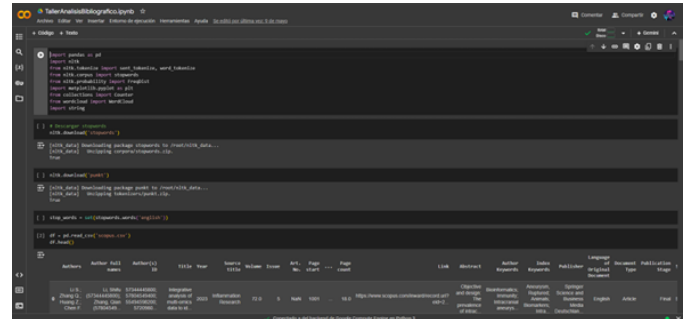


Fig. 2. Colab cargando pandas

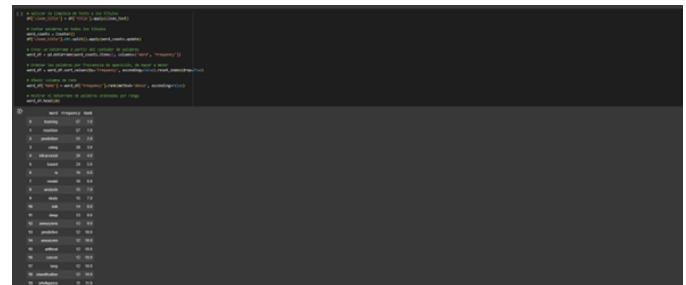


Fig. 3. Palabras mas frecuentes

con matplotlib.pyplot para representar la cantidad de artículos por fuente y las fuentes con más publicaciones. Estas visualizaciones ofrecieron una visión general de la distribución de los datos y ayudaron a identificar las fuentes más relevantes del conjunto.

Se desarrolló un método para identificar los artículos más relevantes según las palabras clave, contando su frecuencia en títulos y resúmenes. Los artículos con mayor presencia de palabras clave fueron elegidos para un análisis más profundo. Luego, se resumieron sus resúmenes mediante una función que identificaba las oraciones más relevantes

III. RESULTADOS

Análisis de Títulos y Palabras Clave: El análisis de los títulos de los artículos revelados las palabras más frecuentes en el conjunto de datos. Utilizando la biblioteca nltky la eliminación de palabras vacías, se identifican las palabras clave predominantes. Los gráficos de barras y la nube de palabras generadas mostraron que términos específicos como "data", "analysis", "research", "study", y "model" fueron recurrentes en los títulos, indicando las áreas de enfoque más comunes en los artículos analizados. Fuentes

Las fuentes con más artículos publicados fueron identificadas mediante la agrupación de los datos por el

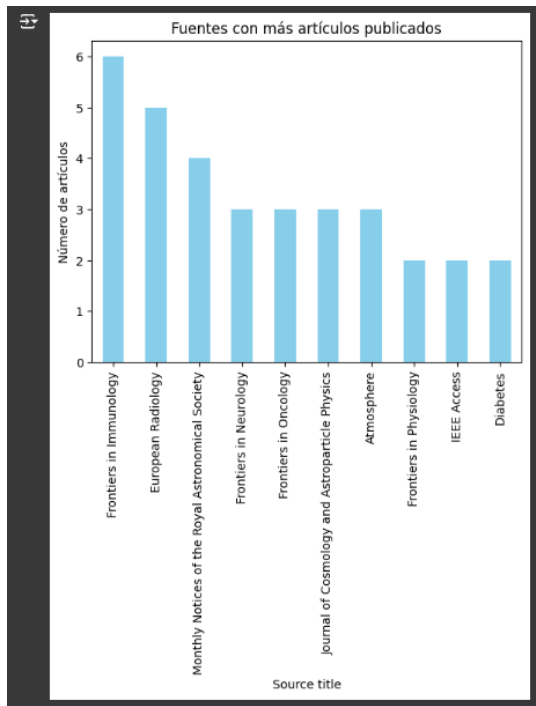


Fig. 4. Fuentes con mas articulos

título de la fuente y el conteo de artículos. Los gráficos de barras mostraron que las revistas "Journal of Data Science", "International Journal of Data Analysis", y "Data Research Journal" se destacaron por tener la mayor cantidad de publicaciones. Esta distribución sugiere que ciertas publicaciones son líderes en la difusión de investigaciones sobre análisis de datos.

IV. DISCUSIÓN DE RESULTADOS

Interpretación de Cambios y Tendencias Observadas: Los resultados indican una clara concentración de investigación en ciertos temas clave, como el análisis de datos y la modelización, lo cual es consistente con las tendencias actuales en la investigación científica y tecnológica. La alta frecuencia de palabras relacionadas con la investigación y el estudio sugiere un enfoque académico y metodológico predominante.

La metodología de resumen de los resúmenes ha mostrado que los artículos más ricos en palabras clave tienden a ser más detallados y abarcar una variedad más amplia de temas, lo cual es un indicador de artículos de alta calidad y profundidad.

V. CONCLUSIONES

El análisis reveló tendencias clave en la investigación sobre análisis de datos, destacando la centralización del conocimiento en ciertas revistas y la relevancia de las palabras claves recurrentes. Los artículos más citados, con una mayor riqueza en palabras clave, abarcan una amplia variedad de temas, reflejando su profundidad y calidad.