

Integrantes Grupo 4

Kendra Artavia Caballero

Alexia Alvarado Alfaro

Jose Soto Perez

Johey Artavia Barrantes

Trabajo en equipo, aplicación de Python en el Análisis de Datos.

1. [1]P. Bhardwaj, C. Choudhury, and P. Batra, “Automating Data Analysis with Python: A Comparative Study of Popular Libraries and their Application,” in 2023 3rd International Conference on Technological Advancements in Computational Sciences (ICTACS), Tashkent, Uzbekistan: IEEE, Nov. 2023, pp. 1243–1248. doi: 10.1109/ICTACS59847.2023.10390032.

Investigación realizada en el Departamento de CSE, Universidad de Chandigarh, Mohali, Punjab, India. El estudio muestra que el nivel de experiencia del usuario y el tipo de tarea de análisis de datos afectan la elección de la biblioteca. Es decir que al elegir la biblioteca adecuada es crucial para automatizar el análisis de datos utilizando Python. El estudio ofrece información detallada sobre las fortalezas y debilidades de las bibliotecas Python conocidas y populares para automatizar trabajos de análisis de datos. Y también los hallazgos del estudio pueden ayudar a las personas a tomar decisiones informadas en una variedad de industrias, incluidas las empresas, la atención médica y las finanzas. Este estudio examinó la eficiencia y efectividad de seis bibliotecas de Python conocidas para trabajos de automatización de análisis de datos. Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn y TensorFlow son las bibliotecas que se analizaron en función de métricas como eficiencia, facilidad de uso y exactitud.

2. [2]D. R. Raman, S. Jayalakshmi, K. Arumugam, A. V. A. Raj, D. Balaji, and R. Brightsingh, “Implementation of Data Analysis and Document Summarization in Social Media Data Using R and Python,” in 2022 4th International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India: IEEE, Sep. 2022, pp. 1457–1464. doi: 10.1109/ICIRCA54612.2022.9985479.

Investigación realizada en el Department of Computer Science Engg, IFET College of Engineering, Villupuram, India. El estudio se basa en un método integral para analizar y resumir los resultados de sitios web de redes sociales que pueden enviar sus flujos de datos a sus respectivas nubes. El enfoque sugerido detectaría valores atípicos en los flujos de datos y podría probarse secuencialmente utilizando programación R y análisis de datos. El estudio logra clasificar el material infringido de las redes sociales y sus mensajeros de una manera efectiva para enseñar a estudiantes y profesionales cómo comportarse de manera ética en las

plataformas sociales. Los datos estadísticos se examinaron utilizando el lenguaje de programación Python.

3. [3]S. B. Sonu and A. Suyampulingam, “Linear Regression Based Air Quality Data Analysis and Prediction using Python,” in 2021 IEEE Madras Section Conference (MASCON), Chennai, India: IEEE, Aug. 2021, pp. 1–7. doi: 10.1109/MASCON51689.2021.9563432.

Investigación realizada en el Department of Electrical and Electronics Engineering, Amrita School of Engineering, Coimbatore Amrita Vishwa Vidyapeetham, India. En este artículo, el análisis de la calidad del aire, llevado a cabo la plataforma Python, implementa un análisis de datos históricos de calidad del aire de la ciudad de Coimbatore y la predicción de los datos de calidad del aire mediante un modelo de regresión lineal. Se analizan los datos, se calcula y visualiza la proporción del principal contaminante PM_{2,5}. Utilizando el modelo de regresión, el conjunto de datos se utiliza para entrenar y probar el modelo en el que se ejecuta la limpieza, el preprocesamiento, el entrenamiento y las pruebas de datos. Se utilizan varias bibliotecas incorporadas en la plataforma Python para la parte de predicción y análisis de datos.

4. [4]G. Yu, “Financial data analysis and risk quantification based on Python,” in 2021 International Conference on Computer, Blockchain and Financial Development (CBFD), Nanjing, China: IEEE, 2021, pp. 214–217. doi: 10.1109/CBFD52659.2021.00049.

Investigación realizada en la Universidad Wenzhou-Kean, Wenzhou, China. El estudio se basa en poder realizar un análisis de los datos financieros como la gestión de riesgos y la inversión cuantitativa. Con el objetivo de un análisis cuantitativo de los riesgos. Se logró determinar como una estrategia de búsqueda de los datos mediante un URL , la aplicación de python también fue utilizada para hacer diversos análisis cuantitativos como el método de varianza que ayuda a ver los riesgos.

5. [5]J. DSouza and S. Velan S., “Using Exploratory Data Analysis for Generating Inferences on the Correlation of COVID-19 cases,” in 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kharagpur, India: IEEE, 2020, pp. 1–6. doi: 10.1109/ICCCNT49239.2020.9225621.

Investigación realizada en el Departamento de Ingeniería y Ciencias de la Computación, Amity University Dubai, Dubai, EAU. El objetivo de este artículo fue obtener conocimiento sobre los efectos de la pandemia con respecto a las variables/etiquetas proporcionadas en el conjunto de datos. En él podemos ver cómo se logró presentar un análisis de los datos para

formular patrones relacionados con la actual pandemia COVID-19 y obtener información de importante de datos. El uso de Python en esta investigación fue mediante el uso de gráficos para generar una mejor visualización de los datos, visualización basadas en las bibliotecas de análisis de datos de Python NumPy, Pandas, Matplotlib, Seaborn, ggplot, Plotly, PyQtGraph, VisPy, Bokeh, Altair y Pygal

6. [6]X. Zhang, J. Huang, Y. Yang, X. He, R. Liu, and N. Zhong, “Applying Python in Brain Science Education,” in 2019 International Joint Conference on Information, Media and Engineering (IJCIME), Osaka, Japan: IEEE, 2019, pp. 396–400. doi: 10.1109/IJCIME49369.2019.00086.

Investigación realizada en Osaka, Japón. El objetivo del estudio fue demostrar el uso de técnicas de neuroimagen en educación en ciencias del cerebro, específicamente en la manipulación y análisis de datos de neuroimagen. El artículo logró presentar casos de enseñanza y práctica con herramientas relevantes basadas en el lenguaje de programación Python y su ecosistema tecnológico, con el fin de ayudar a los estudiantes a adquirir habilidades fundamentales para la manipulación y análisis de datos de neuroimagen. Además, busca que los estudiantes utilicen tecnologías relacionadas con el ecosistema de Python para dominar técnicas de inteligencia artificial como clustering K-means, clustering Ward, PCA, ICA, análisis univariado a gran escala y otras tecnologías relacionadas con Python utilizadas en educación en ciencias del cerebro. Se aplica python en:

Neuroimaging Preprocessing With Python:

- ❖ fMRI Manipulation with Python
- ❖ fMRI Data Normalization with Python

Neuroimaging Analysis with Python:

- ❖ Functional Connectivity Analysis with Python
- ❖ Multivoxel Pattern Analysis with Python

7. [7]P. Gupta, M. Gupta, R. Kansal, L. Singh, M. Khotani, and V. Gupta, “Analysis of Twitter data to Identify Intuitive Mental Health Well Being,” in 2021 5th International Conference on Information Systems and Computer Networks (ISCON), Mathura, India: IEEE, Oct. 2021, pp. 1–4. doi: 10.1109/ISCON52037.2021.9702319.

Se usaron Tweets de personas, por ende el estudio engloba muchos países, sin embargo se publicó en la conferencia 5ta Conferencia Internacional sobre Sistemas de Información y Redes de Computadores (ISCON), Mathura, India. Se propone analizar la salud mental de los usuarios de Twitter utilizando los tweets de los usuarios como nuestra entrada, y al analizar los tweets de las personas generamos la polaridad de los tweets, lo que resulta en el resultado de la salud mental intuitiva. análisis. Este artículo describe una mente sana en general que es necesaria para el bienestar físico y mental de las personas, proporciona un modelo para un análisis de sentimientos que detecta pensamientos positivos alentadores, puntos de vista negativos y pensamientos neutrales, así como varios usuarios, basándose en la técnica del

análisis sentimental. El trabajo incluye preprocesar y segregar datos mediante el clasificador Naive Bayes, y luego encontrar las vistas a través de la polaridad y la subjetividad. Estos sentimientos muestran la salud mental intuitiva de la persona además, sugiere el uso del análisis de sentimientos para analizar la salud mental de las personas.

Con lo obtenido se puede analizar que si la proporción de polaridad negativa o positiva de los tweets es mayor, según la técnica de análisis sentimental el usuario podría tener algunos problemas de salud mental. Entonces, a través del análisis, se intenta hacer un análisis de la salud mental intuitiva de las personas y brindarles alguna ayuda. Se utilizaron diferentes bibliotecas, por ejemplo: textblob, matplotlib, wordcloud, pandas, NumPy. Matplotlib es una biblioteca que se utiliza para trazar gráficos 2D. Numpy y Pandas son las dos bibliotecas de Python que se utilizan para cálculos matemáticos. Wordcloud es la biblioteca utilizada para la visualización de datos según la frecuencia de las palabras. A partir de estos se crearon diferentes gráficos y figuras, entre ellos:

- Gráfico disperso de nuestros hallazgos a través del análisis de sentimiento
- Gráfico de barras que representa la polaridad del sentimiento con el recuento
- La nube de palabras se utiliza para las palabras polarizadas negativas.

8. [8]I. Grout, “Pulse-Frequency Modulation Signal Generation for Programmable Logic Using Python and VHDL,” in 2018 International Electrical Engineering Congress (iEECON), Krabi, Thailand: IEEE, 2018, pp. 1–4. doi: 10.1109/IEECON.2018.8712177.

Investigación realizada en Krabi, Tailandia. El estudio tiene como objetivo, realizar una implementación de circuito en hardware digital utilizando un dispositivo de lógica programable (PLD), específicamente el dispositivo Xilinx CoolRunner-II XC2C256 CPLD. Así logra desarrollar una caja de herramientas para la generación de señales PFM, que permite la generación automática de descripciones de circuitos VHDL para la implementación en CPLD. Se utiliza Python para configurar el proyecto y generar automáticamente el código VHDL necesario para la implementación en un dispositivo CPLD, proporcionando así una solución práctica para la generación de señales analógicas a partir de hardware digital.

9. [9]C. Wang, S. Vemula, and M. Frye, “Out-of-school Time STEM: Teach Programming Using Python for High School Girls,” in 2020 IEEE Integrated STEM Education Conference (ISEC), Princeton, NJ, USA: IEEE, Aug. 2020, pp. 1–6. doi: 10.1109/ISEC49744.2020.9397812.

Investigación realizada en Princeton, NJ, USA. El objetivo del estudio es abordar los desafíos asociados con la enseñanza de programación en entornos fuera del horario escolar (OST, por sus siglas en inglés), específicamente dirigido a estudiantes de secundaria. Se destaca la importancia de mejorar la participación y el aprendizaje en ciencias, tecnología, ingeniería y matemáticas (STEM) para preparar a los estudiantes para la sociedad futura, que estará cada vez más centrada en la tecnología. Además, se enfoca en la identificación temprana del

talento STEM para satisfacer las demandas de la fuerza laboral STEM en los Estados Unidos. El estudio identifica la existencia de numerosos programas STEM fuera del horario escolar y destaca el crecimiento de estos programas con un enfoque particular en la enseñanza de programación mediante el uso de plataformas como LEGO Mindstorms. Se reconoce la necesidad de abordar desafíos específicos en la enseñanza de programación, tanto desde una perspectiva técnica (datos, instrucciones y sintaxis) como desde una perspectiva pedagógica (motivación y problemas técnicos). El estudio aboga por una secuencia efectiva de enseñanza que ofrezca un lenguaje simple y aborde diversos problemas para resolver. Python se utilizó como el lenguaje de programación principal para diseñar e implementar un curso de programación dirigido a estudiantes de secundaria, específicamente a chicas de secundaria.

10. [10]F. Liu and D. Liu, "Mining the Mental Health Crisis Among College Students Using Big Data : ---- A Data Analysis based on TikTok," in 2022 3rd International Conference on Education, Knowledge and Information Management (ICEKIM), Harbin, China: IEEE, 2022, pp. 859–862. doi: 10.1109/ICEKIM55072.2022.00188.

Investigación realizada en China, enfatizado en los estudiantes de este país. El estudio busca comprender mejor el estado del estudio sobre las crisis de salud mental entre estudiantes universitarios en China y buscar nuevos métodos para mejorar su salud mental en la era de la información. Método A través del análisis visual de los grandes datos de TikTo. Con la ayuda de la recopilación de datos y el uso de python para crear tablas, gráficos y figuras se pudo obtener diferentes resultados, entre ellos se destacan: El mapa de partición de videos muestra que además de los resultados de la vida diaria y el estudio en el campus, TikTok también ofrece contenidos sociales, legales y psicológicos que también son resultados importantes. Los comentarios hechos por los estudiantes universitarios en la plataforma Douyin son en su mayoría palabras negativas como "sin palabras", "hurtándose la nariz", "recibido, por favor responda", lo que refleja la tendencia negativa de la mentalidad de los estudiantes universitarios en la actualidad. Se utilizó Software: Rstudio4.1.1, python1.9, gephi, ROSTCM6 Algoritmos: snowNLP, LDA, análisis de redes sociales snowNLP: SnowNLP es una biblioteca de clases escrita en Python, que puede manejar fácilmente contenido de texto chino. Y con la ayuda de este lenguaje se crearon gráficos, tablas, histogramas entre otros para poder tener resultados concisos y poder hacer comparativas.