ASSADOUR KHANDJIAN

+961 79110098 | Armenia Street, Bourj Hammoud, Vartanian Building, Lebanon
| asso796@gmail.com | linkedin.com/in/assadour-khandjian-b1a305119

PROFESSIONAL OVERVIEW

I am a data science and machine learning enthusiast with good skills in Python and different machine learning tools. Natural language processing, data wrangling, statistical modeling and predictive modeling are some of the skills I possess from my hands-on experience. Complexity excites me, and I love using data analysis techniques to provide practical recommendations. Retaining a fanatical passion for perfection, I am all set to introduce groundbreaking solutions under the disguise of formidable data science projects.

WORK EXPERIENCE

Data Scientist/Developer | Softech , Beirut-Lebanon April 2023 - Present

- Prepare Power BI dashboards for Analysis purposes.
- Integrate Power BI and Microsoft Fabric for better dynamic analysis.
- Use Microsoft Fabric to clean, transform and model the data.
- Big data cleaning and processing using Python and Apache Solr.
- Created a tool using Google AI Studio, and saved time to process some of the data that needed to be handled in real time.
- Created tools to integrate with Apache Solr using HTTP requests (Adding Fields, Editing the Data, Creating cores.)
- Managed and structured data (deleting duplicates, handling variable types).
- Leveraged SQL for efficient querying and transformation of structured data, enhancing the overall data processing workflow.
- Ensured data accuracy and maintained a high level of attention to detail while documenting findings and generating reports.
- Successfully managed and analyzed large datasets using Python, identifying key trends and insights crucial for decision-making processes.
- Sentiment analysis on huge Twitter data.
- Analyzing user tweets and counting the most common words used by users.
- Implemented clustering techniques to identify patterns and groupings within large datasets, enhancing data interpretation and decision-making processes.

Data Analyst |Confidential , Beirut-Lebanon December 2010 - Present

- Conducted in-depth analysis of judicial data, applying statistical methods and data visualization techniques to identify patterns, trends, and anomalies, using methods in python.
- Played a pivotal role in anomaly detection and reporting within the Passport system, ensuring the accuracy and integrity of sensitive data.
- Applied data mining techniques to extract valuable information from complex datasets, facilitating more effective data-driven strategies.
- Applied SQL-based data queries to extract valuable information from complex datasets, and fixing any errors in the data that need to be fixed.

EDUCATION

Professional Diploma in Artificial Intelligence and Data Science | American University of Beirut | NLP, Machine Learning, Deep Learning, Data Science | Beirut | Jan 2024

BS in Computer Science | Arts and Sciences University in Lebanon | Beirut | July 2014

S K I L L S

Programming: Python,SQL | **Big Data Technologies:** Apache Spark,Apache Solr | **Data Analysis**: Pandas, NumPy | **Machine Learning:** Scikit-learn, TensorFlow, Keras | **Data Visualization**: Matplotlib, Seaborn,Power | **Statistical Analysis**: Hypothesis Testing and Regression analysis | **Tools**: Jupyter Notebook, Amazon SageMaker

P R O J E C T S:

1. Heart Disease Diagnosis Using a Machine Learning Algorithm

- Project Focus: Predicting heart disease using data mining techniques.
- Importance: Heart disease is a leading global cause of mortality, emphasizing the need for effective identification and management.
- Methodologies Explored: Association rules, classification, clustering, and prediction through data mining.
- Utilized Machine Learning Algorithms: Logistic Regression, Random Forest, Naïve Bayes for accurate predictions.
- Dataset Source: UC Irvine Machine Learning Repository.
- Data Processing and Visualization: Gain insights into factors influencing heart disease, including age, gender, chest pain type, blood pressure, cholesterol levels, and other health variables.
- Apache Spark Integration: Suggested using Apache Spark for efficient processing of large datasets and real-time analysis.
- Results : Random Forest stood out, achieving an impressive accuracy rate of 81%.

2. Twitter Data Sentiment Analysis Using Python and Apache Spark

- Project Focus: Conducting sentiment analysis on Twitter data using Python and Apache Spark.
- Implementation of Two Systems: One using Apache Spark with distributed computing and MLlib. Another using Python with NLTK and Scikit Learn.
- Key Steps in Sentiment Analysis:
    - Data Preprocessing: Cleaning and preparing data (remove RT, #, web links).Eliminating stop words and removing punctuation marks.
    - Sentiment Score: Calculating sentiment polarity with the textblob library (values between -1 and 1).
    - Feature Extraction: Utilizing TF-IDF for numerical representation of tweets.
    - Machine Learning Modeling: Choosing logistic regression as the sentiment classification algorithm. Training and testing both systems using preprocessed data and feature vectors.
    - Results: Apache Spark outperforms Python in terms of training time, accuracy, and F1-score.
- Future Work: Incorporating real-time features using the Twitter API for analyzing sentiments from millions of users.