



SYED MUJTABA ASHAR

DATA SCIENTIST

 +92-3217440028

 syedasharj@gmail.com

 syed-ashar-7b4693302

 <https://github.com/SyedMujtabaAshar>

EDUCATION

Master in Data Science

FASTNUCES , Pakistan
2024 ~ Ongoing

Bachelors of Science in Physics

COMSATS University, Pakistan
2019-2023

EXPERTISE

- Concepts of OOP
- C++ | Python
- Pandas | Numpy
- Matplotlib | SeaBorn
- MS Word | MS Excel
- Power Bi
- Data Base : SQL
- Data Structures (C++)
- MATLAB
- RadiAnt DICOM viewer
- Fortran

PROFILE

I am currently pursuing a Master's degree in Data Science with a strong foundation in statistics and mathematics. I have solid expertise in Python, C++, and SQL, along with a deep understanding of object-oriented programming, data structures and algorithms, and database management systems. Known for my analytical mindset and collaborative approach, I bring both creativity and precision to data-driven problem-solving.

WORK EXPERIENCE

Spaltalink (remote)

March - Sep 2023

Junior Data Analyst ~ Internship

- Conducted data cleaning, transformation, and preparation to ensure datasets were accurate, consistent, and ready for analysis.
- Created reports and visualizations to communicate insights effectively, aiding in decision-making processes for various projects.

SUPARCO Islamabad (On-Site)

June - August 2024

Internship

- Gathered and analyzed large datasets to identify trends and patterns, supporting strategic initiatives within the organization.

PROJECTS

- Chat With your PDF's using Langchain (LLM)
- Library Management System (OOP)
- Data Cleaning And Visualization
- House Pricing Prediction
- Sentiment Analysis
- Store Progress Dashboard Via Power Bi
- Online Book Store (DBMS)
- COVID-19 Report of India Via Power Bi
- Project for Handling Missing values in Dataset

Languages Use

- Python
- C++
- R

IDE (Integrated Development Environment)

- *Visual Studio Code*
- *Jupyter Notebook*
- *Google Colab*

Libraries:

- *Pandas*
- *Matplotlib*
- *Seaborn*
- *Numpy*
- *Streamlit*
- *Flask*
- *scikit-learn*

PROJECTS DETAILS

1. Chat With your PDF's using Langchain (LLM):

Features:

- **Question and Answer:** Answer questions based on the content of the uploaded PDF.
- **Summarize PDF:** Generate a summary of the PDF document.
- **Translate PDF:** Translate the content of the PDF into another language.
- **Generate Questions:** Create meaningful questions from the text in the PDF.

2. Library Management System (OOP):

Features:

- **Multi-Class Sentiment Classification:** Classifies reviews as Positive, Negative, or Neutral.
- **Text Preprocessing:** Cleans and prepares data by removing special characters and stopwords.
- **Real-Time Predictions:** Allows users to input reviews and receive immediate sentiment classifications.
- **Improvement Potential:** Achieves ~65% accuracy, with opportunities for enhancement through tuning and alternative algorithms.

3. Data Cleaning And Visualization:

Features:

- **Data Source:** Utilizes the Titanic passenger dataset from Kaggle, containing attributes related to each passenger.
- **Survival Rate Analysis:** Analyzes survival rates based on factors such as passenger class, age, and family size.
- **Insights Extraction:** Extracts meaningful insights to understand the impact of various factors on survival outcomes.
- **Visualization:** Employs data visualization techniques to present findings and trends effectively.

4. House Pricing Prediction:

Features:

- **Data Cleaning:** Prepares the dataset by filling in missing values and converting text categories into numbers for analysis.
- **Two Prediction Models:** Uses two different methods (Linear Regression and Random Forest) to predict house prices.
- **Model Accuracy Check:** Measures how well the models perform by calculating errors and scores to see how close the predictions are to actual prices.
- **Visual Results:** Creates easy-to-understand plots that show how predicted prices compare to real prices, helping to visualize the model's accuracy.

Languages Use

- Python
- C++
- R

IDE (Integrated Development Environment)

- *Visual Studio Code*
- *Jupyter Notebook*
- *Google Colab*

Libraries:

- *Pandas*
- *Matplotlib*
- *Seaborn*
- *Numpy*
- *Streamlit*
- *Flask*
- *scikit-learn*

5. Sentiment Analysis :

Features:

- **Text Cleaning:** The project cleans up movie reviews by removing unnecessary characters and common words that don't add meaning, making the text easier to analyze.
- **Sentiment Classification:** It can tell if a review is Positive, Negative, or Neutral based on the words used in the text.
- **Real-Time Predictions:** Users can type in a movie review, and the model will quickly give back a sentiment prediction, like "Positive" or "Negative."
- **Improvement Opportunities:** The model currently has an accuracy of about 65%, and there are ways to make it better, such as trying different algorithms or improving how the text is processed.

6. Employee Attrition Prediction :

Features:

- **Attrition Prediction:** The project uses machine learning to predict if an employee is likely to leave the company based on various HR-related factors.
- **Data Analysis:** It includes an exploratory data analysis (EDA) to identify important factors that influence why employees leave, helping to understand the underlying issues.
- **Model Training:** The project trains classification models, specifically Random Forest and Logistic Regression, to evaluate their effectiveness in predicting employee attrition.
- **Actionable Insights:** It provides insights for HR teams to develop strategies aimed at improving employee retention and reducing turnover rates.

7. Store Progress Dashboard Via Power Bi :

Features:

- **Sales Distribution:** Looks at how sales are spread out, showing that most sales are on the lower end, which suggests there's room to grow.
- **Seasonal Trends:** Finds that sales go up during holiday seasons, helping businesses plan for busy times.
- **Top Products:** Identifies the top 5 products that make the most money, so businesses know which items to focus on.
- **Regional Performance:** Analyzes sales by location, showing that places like New York and Los Angeles have the highest sales potential, guiding where to target efforts.

8. ANN For Breast Cancer Prediction :

Features:

- Developed an Artificial Neural Network to predict breast cancer using diagnostic data.
- Achieved high frequency with well optimized architecture