Muskan Ara

muskan17ara@gmail.com | linkedin.com/in/muskan-ara-53bb192a8 | https://github.com/Muskan896F

About Me

"I'm an aspiring Data Scientist with experience in Python, machine learning, and data analysis. I recently completed a one-month online internship at Cognifyz, where I worked on data cleaning, analysis, and visualization. I'm eager to apply my skills in a role that allows me to contribute to data-driven decision-making and continuous learning. I'm passionate about using data to solve problems and provide insights."

Experience

Data Science Intern   *Sep–Oct 2024*
Cognifyz, Remote

- Conducted data cleaning and preprocessing to ensure the quality and reliability of datasets for analysis.
- Utilized Python for data analysis and visualization, leveraging libraries such as **pandas**, **NumPy**, and **Matplotlib**.
- Designed and developed machine learning models for predictive analysis, evaluating their performance to generate actionable insights.
- Gained expertise in handling, analyzing, and building machine learning models for real-world applications, improving decision-making processes.

Data Science Intern   *Dec 2024–Jan 2025*
Oasis Info Byte, Remote

- Leveraged advanced NLP techniques, such as lemmatization, stemming, and **word2vec**, to identify duplicate questions and improve user experience on the Quora platform.
- Implemented the **Random Forest** algorithm for classification tasks and integrated **Streamlit** to create an intuitive user interface for real-time interaction.
- Developed a machine learning-based system to identify defective semiconductor wafers, optimizing production efficiency and maintaining high-quality standards in integrated circuits.
- Applied advanced machine learning techniques to enhance defect detection accuracy in semiconductor manufacturing, improving operational efficiency.

Skills

- **Programming Languages:** Python, SQL, MongoDB
- **Data Analysis:** NumPy, pandas, Matplotlib, seaborn
- **Machine Learning:** Supervised Learning, Unsupervised Learning, Recommendation Systems, NLP techniques (lemmatization, TF-IDF, word2vec)
- **Deep Learning:** ANN, CNN, RNN, LSTM, GRU, Transformers
- **Libraries:** scikit-learn, TensorFlow, Keras, NLTK, Gensim, spacy
- **Web API Development:** Flask, Streamlit, FastApi
- **Tools & Technologies:** VS Code, Jupyter Notebook, Git, GitHub, Docker, MongoDB ,MySQL Workbench
- **Other Skills:** Model deployment, working with APIs, End-to-End project handling

Projects

- **Reducing Operational Costs with Cost-Sensitive XGBoost-Based Failure Prediction** *GitHub Link*
  - This project focuses on handling imbalanced sensor data for classification tasks, with techniques such as data preprocessing, model evaluation, and cost analysis.
  - The dataset contains sensor data with missing values, class imbalance, and various features requiring imputation and scaling.
  - Several machine learning models, including Random Forest, XGBoost, and CatBoost, are evaluated using metrics such as accuracy, F1-score, precision, recall, and ROC-AUC score.
  - Advanced techniques such as KNN imputation, SMOTE-TOMEK for resampling, and robust scaling are used to improve model performance and handle data irregularities.
  - The project analyzes multiple experiments, comparing models' performance and costs, with the goal of identifying the most effective solution for imbalanced datasets.

- **AI-Driven Pneumonia Diagnosis: Harnessing Custom CNNs for Chest X-Ray Analysis** *GitHub Link*
  - Developed a custom convolutional neural network (CNN) to detect pneumonia from chest X-ray images.
  - Achieved high diagnostic accuracy through rigorous training and validation processes.
  - Deployed the model in a user-friendly web application for healthcare professionals.

- **Building E-Commerce Products Recommendations using Machine Learning and Flask** *GitHub Link*
  - Implemented a recommendation system using collaborative filtering techniques for an e-commerce platform.
  - Built a Flask-based web interface to deliver personalized product suggestions to users.
  - Optimized the recommendation pipeline to handle large-scale data efficiently.

- **Develop an Advanced Resume Screening App Using NLP and Python** *GitHub Link*
  - Created an advanced resume screening app leveraging NLP techniques to match resumes with job descriptions.
  - Utilized libraries like spaCy and Gensim to extract and compare relevant skills.
  - Integrated ChromaDB for efficient skill matching and a seamless user experience.

- **End-to-End Machine Learning Pipeline for Zomato Delivery Time Predictions** *GitHub Link*
  - Designed an end-to-end machine learning pipeline to predict Zomato delivery times based on various factors.
  - Conducted exploratory data analysis and feature selection to identify key influencers of delivery times.
  - Deployed the pipeline on a cloud platform for real-time predictions and insights.

Certifications

- **Google Data Analytics** — Coursera
  October 2024
  Focused on data analysis techniques, data visualization, and cleaning.

- **Machine Learning and Deep Learning** — Udemy
  July 2024
  Covered algorithms, neural networks, and deep learning frameworks.

- **Complete GenAI Course with LangChain and HuggingFace** — Udemy
  September 2024
  Specialization in Generative AI, LangChain integration, and HuggingFace models.

Education

Master of Science in Data Science 2023 - 2025
Chandigarh University, Chandigarh, India
Key Courses: Machine Learning, Data Visualization, Big Data Analytics, Statistical Modeling
Highlights: Academic projects focused on predictive analytics and real-world data solutions.