Mahboob Alam

6396281660 mahboobalam7131@gmail.com https://www.linkedin.com/in/mahboobalam786/
https://github.com/MahboobAlam0

Introduction

My name is Mahboob Alam. I am pursuing M.Tech in Data Science. I am eager to work in a collaborative environment where I can learn from experienced professionals, contribute to impactful projects, and grow both technically and analytically. I am particularly interested in exploring large datasets, uncovering insights, and using data to support informed decision-making.

I am proficient in Python, SQL, and data visualization, enabling me to analyze complex datasets effectively. Additionally, I have a keen interest in machine learning and statistical modeling, which help in making accurate predictions and optimizing business strategies.

Projects

Movie Recommender System | *Cosine Similarity, Streamlit*

- Built a content-based movie recommender system using the TMDB Kaggle dataset.
- Preprocessed data by extracting features such as genres, keywords, and movie descriptions, and combined them into a unified feature set.
- Implemented **cosine similarity** to measure the similarity between movies and recommend top-N similar movies based on a given input.
- Improved recommendations by incorporating movie metadata and optimizing the feature engineering process.
  Developed a user-friendly function to retrieve similar movies, tested with popular titles, and validated results.

Customer Churn Predictive Analysis | *RandomForestClassifier, Streamlit*

- Developed a RandomForestClassifier model to predict customer churn for targeted marketing.
- Analyzed customer behavior using EDA and feature engineering.
- Improved model accuracy through data preprocessing, handling imbalanced data, and hyperparameter tuning.
- Evaluated performance using precision, recall, F1-score, and AUC-ROC.
- Implemented in Python (Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn).
- **Live Demo:** https://customerchurn-0.streamlit.app/

Medical Insurance Cost Analysis Predictor | *Multiple Regression, Streamlit*

- Developed a predictive model using Linear Regression to estimate medical insurance costs.
- Achieved **78.30%** accuracy in predictions by analyzing key factors such as age, BMI, smoking status, and region.
- Performed **Exploratory Data Analysis (EDA)** to understand feature importance and data distribution.
- Applied data preprocessing techniques like handling missing values, encoding categorical variables, and scaling numerical features.
- Evaluated model performance using **R^2 score, MSE, and RMSE** to optimize accuracy.
- Implemented the project in Python using **Pandas, NumPy, Matplotlib, Scikit-learn** for model training and evaluation.
- **Live Demo:** https://medicalinsurancecostanalysis0.streamlit.app/

Technical Skills

Languages: Python, C
Technologies: Numpy, Pandas, Matplotlib, Seaborn, Scikit-learn, Machine Learning
Databases: MySQL.
Tools: VSCode, Jupyter Notebook, Google Colab.
Operating Systems: Windows, Linux, MacOS.

Education

Defence Institute of Advanced Technology – DRDO, Pune 2024-2026
Master of Technology degree with a specialization in Data Science (7.31SGPA)

Rao Birender Singh State Institute of Engineering and Technology, Rewari 2019-2023
Bachelor of Technology degree with a specialization in Computer Science and Engineering (58.29 %)

Bal Bharti Inter College, Amroha 2017-2018
Intermediate with PCM (73.8 %)