Arka Barua

+880 1864-644612 arka.barua.173@gmail.com Chattogram, Bangladesh
arkabarua173.github.io/ark Arka Barua ArkaBarua173

Summary

As a recent Computer Science and Engineering graduate seeking opportunities as a Data Analyst and Machine Learning Intern, I bring a solid foundation in programming languages and a keen interest in data analysis and machine learning. Proficient in Python, Pandas, Matplotlib, Seaborn, Plotly, scikit-learn, and eager to apply my skills in real-world projects.

Education

BSC North South University, Computer Science & Engineering 2017-2023

- **CGPA:** 3.22/4.0
- **Trail:** Artificial Intelligence and Networks
- **Coursework:** Design and Analysis of Algorithms, Database Systems, Machine Learning, Pattern Recognition.

HSC Chittagong Govt. Model School & College, Science 2016

- **GPA:** 3.93/5.0

SSC Chittagong Govt. High School, Science 2014

- **GPA:** 5.0/5.0

Certificate Courses

Machine Learning A-Z: AI, Python & R + ChatGPT Prize [2024] May. 2024
(Issued By Udemy)
www.udemy.com/certificate/UC-973b796e-f7bf-4729-9ca6-3e96d4301eac

Advanced SQL May. 2024
(Issued By Kaggle)
www.kaggle.com/learn/certification/arkabarua173/advanced-sql

Skills

Languages: C, C++, Python, PHP, HTML, CSS, SQL, JavaScript

Libraries: Pandas, Scikit-learn, Matplotlib, Plotly, React, Tailwind CSS

Software: Visual Studio Code, PowerBI, Excel, Microsoft SQL Server, PostgreSQL

Projects

Binary Classification with an Imbalanced Dataset | Feb. 2024

- This project aims to predict the probabilities of customers exiting the bank.
- XGBoost is used to address the imbalance by increasing the weights of the minority class.
- The model was evaluated using accuracy, precision, F1 score, recall, and AUC-ROC score.
- SHAP (Shapley Additive Explanations) is used to understand the model.
- **Tools Used:** Python, Pandas, Plotly, Scikit-learn, XGBoost, SHAP

Abalone Age Regression with Ensemble Method | April 2024

- This project aims to predict the age of Abalone.
- XGBRegressor, CatBoostRegressor and LGBMRegressor are trained.
- A Voting Regressor (XGBRegressor + CatBoostRegressor + LGBMRegressor) with soft voting is trained as the final model.
- The primary evaluation metric used is Root Mean Squared Logarithmic Error.
- **Tools Used:** Python, Pandas, Matplotlib, Seaborn, Scikit-learn, XGBoost, CatBoost, LightGBM.

Vehicle Sales Analysis | April 2024

- This project investigates key insights such as average selling prices across car brands and models, comparison of automatic and manual transmission cars, examination of color distribution's impact on prices, analysis of car condition ratings, and the effect of odometer readings on selling prices.
- **Tools Used:** Python, Pandas, PowerBI.

Clustering of Coffee Quality | May. 2024

- This project aims to cluster coffee quality based on similarity with PCA and K-means.
- A dimensionality reduction technique PCA is used for this dataset while keeping important patterns and trends.
- K-means clustering is then used, with the elbow technique determining that 6 clusters are the ideal number.
- **Tools Used:** Python, Pandas, Plotly, Scikit-learn.

References

Sumoy Barua
Lead Software Engineer, Cefalo
Contact: +880 1610-001914

Dr. Mahdy Rahman Chowdhury
Associate Professor, North South University
Contact: +88 02 55668200