Md Mustain Billah

Bangladesh ✉ mustainbillah017@gmail.com ☐ +8801724195527 ⬛ https://www.linkedin.com/in/mdmustainbillah0/

PROFESSIONAL SUMMARY

Data scientist with a focus on **data analytics** and **end-to-end machine learning model implementation**. Aiming to become an expert in leveraging machine learning to enhance real-world business processes. Proficient in data modeling, statistical methods, and predictive analytics using advanced algorithms.

SKILLS

- Python, SQL, R, MLOps, MLflow, DVC, Docker, Git, GitHub, Tableau, MongoDB, MariaDB, Spark.
- Scikit-Learn, NumPy, Pandas, Matplotlib, Seaborn, TensorFlow, Keras, NLTK, Flask, Selenium, Plotly, Streamlit, PyMongo, pySpark etc.
- Machine Learning, Deep Learning, NLP, Time Series, ETS, etc.

EDUCATION

Masters in Applied Statistics and Data Science | Jahangirnagar University | Dhaka, Bangladesh | 2024 | 3.88 CGPA

- **Relevant Coursework:** Introduction to Data Science with Python, Statistical Methods, Probability and Probability Distributions, Machine Learning, Time Series Analysis and Forecasting, Statistical Inference, Data Mining, Big Data, Research Methodology.

Bachelor of Computer Science and Engineering | IUBAT University | Dhaka, Bangladesh | 2022 | 3.58 CGPA

- **Relevant Coursework:** Data Structure and Algorithms, Database Management System, Linear Algebra, Calculus, Software Engineering, Operating System, Data Communication and Computer Network, Business Communication.

DATA SCIENCE EXPERIENCE

Data Analyst | Monico Technologies Ltd. | Dhaka, Bangladesh | August 2023 – Present

- Developed a Python script to extract data from MongoDB using chunking, preventing memory overflow and optimizing runtime by 50%. Implemented Dask for efficient processing of over **500GB** of sensor data, ensuring scalability and expedited analysis.
- Implemented advanced machine learning for autonomous fuel event **classification**, enhancing proactive management. Post data preprocessing, developed business dashboards using Streamlit for streamlined decision-making and operational resilience.
- Led data optimization initiatives, achieving an **84.7%** reduction in data size. Identified and eliminated duplicates in OBD CAN data, resulting in substantial memory savings and fortifying operational efficiency.
- Python, Pandas, Seaborn, Matplotlib, Plotly, Dask, Tableau, Streamlit, pyMongo, MongoDB.

Data Scientist – Internship | Koryntia | London, UK – Remote | May 2023 – June 2023

- Conducted in-depth analysis of diverse **financial datasets** at Koryntia, applying advanced data analysis techniques and statistical methods. Achieved a 15% improvement in risk assessment accuracy and a 20% increase in predicting loan outcomes for users.
- Proficiently utilized data visualization tools to create insightful plots and charts, contributing to informed decision-making processes.
- Machine Learning, Python, Pandas, Numpy, Scipy, Matplotlib, Seaborn, Scikit-learn.

AI/ML Engineer – Intern | Titan Technologies Ltd. | Dhaka, Bangladesh | February 2022 – May 2022

- Innovated and crafted an **HR automation** AI ChatBot, collaborating with product managers and developers to optimize HR processes, leading to a 30% time reduction and significant contribution to company growth.
- Applied statistical analysis, NLP, and machine learning to large datasets for robust feature engineering, selection, modeling, and accurate prediction.
- Collaborated with senior developers on cutting-edge research projects, utilizing advanced technologies such as Selenium, TensorFlow, Keras, and Reinforcement Learning.

PROJECTS

End-to-End Fraud Detection Mastery: Real-Time Machine Learning Excellence | https://github.com/mddmustainbillah/fraud-transaction-detection

- Developed and optimized a real-time fraud detection system using advanced machine learning techniques, online **data ingestion**, and robust **validation**, ensuring high accuracy in identifying fraudulent transactions.
- Implemented a streamlined data preprocessing **pipeline** for optimal model training and conducted comprehensive evaluations, achieving efficiency in fraud detection measured by precision, recall, and F1 score metrics.
- Orchestrated a user-friendly **deployment** through Streamlit, facilitating stakeholder interaction with the model and enabling detailed data analysis, while **automating retraining** for enhanced adaptability to evolving fraud patterns.

Conducted In-Depth Sales Data EDA with Streamlit and Plotly Dashboard | https://github.com/mddmustainbillah/Superstore_plotly_streamlit_dashboard

- Led a thorough **exploratory data analysis** on sales data using Streamlit and Plotly, highlighting regional trends and insights.
- Employed diverse **visualizations**, including bar graphs, tree maps, pie charts, and time series plots, to effectively communicate complex patterns and trends in the sales data.

INVOLVEMENT

International University of Business Agriculture and Technology

- Served as an **IT Mentor** and **Academic Mentor** at the Department of Computer Science and Engineering from November 2019 to April 2022.
- Invited to **speak** at my university on various **machine learning algorithms**, showcasing my passion for the field and the ability to communicate complex technical concepts to diverse audiences of students and faculty.