Md Mustain Billah

Dhaka, Bangladesh | mustainbillah017@gmail.com | +8801724195527 | linkedin.com/in/mdmustainbillah0/
github.com/mddmustainbillah | mddmustainbillah.github.io/ | youtube.com/@mustainbillah3050

Education

Masters in Applied Statistics and Data Science — Jan 2023 - Dec 2023
Jahangirnagar University, Dhaka — *CGPA: 3.88/4.00*

Bachelor of Computer Science and Engineering — Jan 2018 - Apr 2022
IUBAT, Dhaka — *CGPA: 3.58/4.00*

Experience

Data Analyst | **Monico Technologies Ltd.** | **Dhaka, Bangladesh** — Aug 2023 - Dec 2024

- Built **live KPI dashboards** for vehicle statistics, improving UX and **recovering 35% of churned customers.**
- Optimized **MongoDB data extraction** by 20% through chunking, preventing memory issues.
- Implemented **Dask for scalable analysis**, cutting processing time by **25%** for sensor data.
- Reduced data size by 27% through optimization and deduplication for better efficiency.
- Developed **ML-based fuel event classification system** with business dashboards for improved decision-making.

Data Scientist - Intern | **Koryntia** | **London, UK - Remote** — Jun 2023 - Jul 2023

- Developed a **loan prediction model** using **RandomForest**, improving accuracy by 15% through feature engineering.
- Conducted **statistical analysis** of loan default patterns, identifying key risk factors through correlation analysis.

AI/ML Engineer - Intern | **Titan Technologies Ltd.** | **Dhaka, Bangladesh** — Feb 2022 - Jun 2022

- Built and deployed NLP-based HR chatbot, handling 200+ daily employee queries.
- Automated **HR query handling processes**, reducing manual workload by 30%.

Projects

Hybrid Recommendation System - Personalized Music Engine | *github.com/mddmustainbillah/hybrid-recommendation-system*

- **Designed and built** a dynamic hybrid recommender system using collaborative and content-based filtering to deliver personalized music suggestions at scale.
- Increased user **engagement and CTR by 15–30%**, and supported free-to-paid **user conversion** and retention.
- **Reduced data volume by 99.95%** (from 60GB to 30MB) by implementing SciPy sparse matrices and Dask-based **parallel processing**, enabling real-time recommendations across 1M users × 30K songs.
- Solved key production challenges including **cold-start issues**, slow loading, and static user preferences by implementing fallback logic, Streamlit caching, and **dynamic hybrid weighting**, ensuring scalability and user-centric performance.

Real-Time YouTube Sentiment Analysis Engine (MLOps-Driven) | *github.com/mddmustainbillah/yt-sentiment-analysis*

- **Improved model accuracy** from 65% to 87% (**22% absolute increase**) by applying SMOTE, fine-tuning, and stacking LightGBM, Logistic Regression, and KNN.
- **Saved hours of wasted effort** by enabling users to skip low-quality YouTube tutorials, through a real-time sentiment analysis tool built with a **Flask API** and **Chrome extension**.
- **Engineered a reproducible MLOps pipeline** using DVC for data versioning, MLflow for experiment tracking, and Docker + CI/CD for automated model training and deployment.

Interactive Sales Dashboard | *github.com/mddmustainbillah/Superstore*$_{plotly_{streamlit_{dashboard}}}$

- Developed Streamlit-based **dashboard** with Plotly visualizations for **multi-dimensional sales analysis and filtering.**
- Created interactive visualizations to **identify top-performing regional markets.**

Skills

Languages & Databases:	**Python** (Advanced), **SQL**, MongoDB
ML & MLOps:	**Scikit-learn**, TensorFlow, **MLflow, DVC**, Prefect, Pipeline Orchestration
Data Processing & Visualization:	**Pandas**, NumPy, **Dask**, Matplotlib, Seaborn, Plotly, BeautifulSoup
API & DevOps:	**Docker**, CI/CD, GitHub Actions, **FastAPI**, Flask, Streamlit, Postman, Git
Core Competencies:	**Machine Learning**, **MLOps**, Statistical Analysis, Problem-solving, Communication

Teaching & Leadership

- Led **live online Data Science classes** for 50+ students, teaching ML and hands-on Python implementations.
- Mentored **CS students and delivered guest lectures on ML/AI**, fostering technical growth in aspiring developers.

Certifications

SQL | 365 Data Science
Python Programmer Bootcamp | 365 Data Science
Machine Learning in Python | 365 Data Science
Git and GitHub | 365 Data Science
Introduction to Tableau | 365 Data Science
Deep Learning with TensorFlow 2 | 365 Data Science