

NELOY PRAMANIK SUPTO

Data Analyst | Machine Learning Engineer | Applied ML & Data Science

Mirpur, Dhaka, Bangladesh | +880 1795 302949 | neloybbs@gmail.com

<https://linkedin.com/in/neloy-pramanik-supto-b52677236> | <https://github.com/Neloy4321> |

sites.google.com/diu.edu.bd/nps49 | ORCID 0009-0001-5172-870X



PROFESSIONAL SUMMARY

Computer Science and Engineering graduate specializing in machine learning, applied data science, and data analytics. Author of a peer-reviewed conference paper and multiple research manuscripts in applied ML, with hands-on experience building end-to-end ML pipelines in Python using scikit-learn, TensorFlow, XGBoost, and LightGBM. Skilled in data preprocessing, feature engineering, model evaluation, and model explainability (SHAP, LIME), with beginner-level exposure to model deployment using FastAPI, Docker, and AWS. Currently applying data acquisition, cleaning, and validation workflows in a professional data operations role.

TECHNICAL SKILLS

Programming & Databases: Python, SQL, C (basic), MongoDB

Machine Learning & AI: Scikit-learn, TensorFlow, Keras, XGBoost, LightGBM, Deep Learning (CNN, LSTM), NLP, Transformers (BanglaBERT, mBERT), Model Explainability (SHAP, LIME)

Data Analysis & Processing: Pandas, NumPy, SciPy, statsmodels, Feature Engineering, Data Preprocessing, Data Cleaning & Validation, Exploratory Data Analysis, Statistical Testing

Data Visualization: Matplotlib, Power BI, Excel, Power Query

ML Engineering & Deployment (beginner): FastAPI, Docker, AWS (EC2, ECR), GitHub Actions (CI/CD)

Tools & Platforms: Git/GitHub, Visual Studio, Jupyter Notebook, Linux, Node.js, Playwright

Other: Computer Networking fundamentals (CCNA, MTCNA certified)

RESEARCH & PUBLICATIONS

Supto, N. P., Chowdhury, R., Rahman, M. *Addressing Behavioral Patterns of Late Sleepers Using a Supervised Learning Approach*. Proceedings of the International Conference on Intelligent Data Analysis and Applications (IDAA 2025), Atlantis Press, pp. 341–357, 2026. DOI: 10.2991/978-94-6239-664-7_24. **(Published; presented at IDAA 2025.)**

- Applied supervised learning to classify behavioral patterns of late sleepers from structured survey data; performed feature engineering, model training, and comparative performance evaluation.

Supto, N. P., Chowdhury, R., Das, S., & Biswas, S. *Hybrid CNN–LSTM Framework for Long-Term Solar Irradiance Prediction Using Multi-Source Satellite Data: A Spatio-Temporal Analysis over Bangladesh*. SSRN, 2025. DOI: 10.2139/ssrn.5928559. **(Preprint.)**

- Designed a hybrid deep learning architecture combining CNN and LSTM layers for spatio-temporal solar irradiance forecasting from multi-source satellite data.

Supto, N. P. *An Explainable Machine Learning Framework for Personalized Sleep Health Assessment and Behavioral Recommendations*. SSRN, 2025. DOI: 10.2139/ssrn.5936725. **(Preprint; revised version resubmitted.)**

- Built an interpretable ML framework for personalized sleep health assessment, emphasizing explainability and behavioral recommendation generation.

Additional manuscripts under review:

- Reanalysis bias correction of daily agro-climatic data over Bangladesh (NASA POWER, ERA5-Land vs. 36 BMD stations) using tree-based ensemble models (LightGBM) with leave-one-station-out validation and SHAP-based interpretation.
- Comparative study of BanglaBERT vs. multilingual BERT for Bengali fake-news and hate-speech detection, with quantitative explainability (SHAP/LIME faithfulness) and statistical significance testing (McNemar, bootstrap CI).

SELECTED PROJECTS

Sleep Health Prediction System — End-to-End ML Pipeline

Python, scikit-learn, FastAPI, MongoDB, Docker, AWS EC2/ECR, GitHub Actions · [GitHub](#)

- Developed a complete machine learning system to predict sleep-related health conditions from lifestyle, behavioral, and medical features using a self-collected dataset.
- Engineered a modular pipeline covering data ingestion, validation, transformation, feature engineering, multi-label encoding, model training, and prediction.
- Deployed the model as a FastAPI web service, containerized with Docker, and automated build-and-deploy to AWS EC2 via AWS ECR and GitHub Actions CI/CD.

Agro-Climatic Data Reconstruction with Machine Learning (BD-AgroClim)

Python, scikit-learn, XGBoost, LightGBM, SHAP, Pandas, NumPy · [GitHub](#)

- Evaluated NASA POWER and ERA5-Land reanalysis data against 36 meteorological stations and built a residual-learning bias-correction framework using tree-based ensembles.
- Selected LightGBM after model comparison; reduced maximum-temperature RMSE from ~2.25°C to ~1.36°C under strict leave-one-station-out (LOSO) cross-validation.
- Applied SHAP analysis to interpret correction drivers (soil moisture, terrain variability, vapour pressure deficit) and validated spatial transferability on held-out island stations.

Bengali Fake-News & Hate-Speech Detection with Explainable AI

Python, Transformers (BanglaBERT, mBERT), scikit-learn, SHAP, LIME · [GitHub](#)

- Fine-tuned and compared BanglaBERT and multilingual BERT on two Bengali datasets (~95k news articles; ~50k comments) against TF-IDF classical ML baselines.
- Evaluated results with accuracy, macro-F1, and ROC-AUC, and validated differences using McNemar's test and bootstrap confidence intervals.
- Quantified explanation faithfulness (comprehensiveness, sufficiency) and SHAP–LIME agreement to assess interpretability across text lengths.

Geo-Business Data Collection Pipeline

Node.js, Playwright, JavaScript, CSV data processing · [GitHub](#)

- Built a modular, multi-stage web-data pipeline to extract and structure publicly available business information, validated on 10,000+ records across US and UK regions.
- Implemented parallel scraping, duplicate removal, retry handling, and CSV export, achieving ~98% address and ~95%+ phone/region extraction accuracy on manual verification.

EXPERIENCE

Data Engineering Associate (Data Operations & Analytics)

Mar 2026 – Present

Red Digital Marketing Group · Part-time · Remote, Dhaka, Bangladesh

- Conduct data acquisition, preprocessing, and validation workflows supporting marketing analytics and business insights.
- Perform data cleaning and transformation to convert raw, unstructured data into structured, analysis-ready datasets, applying consistency checks and quality-control rules.
- Organize datasets using basic data modeling to improve usability for reporting and downstream analytics.

EDUCATION

B.Sc. in Computer Science and Engineering (CSE)

2021 – 2025

Daffodil International University, Dhaka, Bangladesh · CGPA: 3.60 / 4.00

- Relevant coursework: Data Mining & Machine Learning, Artificial Intelligence, Natural Language Processing, Big Data & IoT, Statistics & Probability, Database Management Systems, Data Structures & Algorithms.

CERTIFICATIONS

- Machine Learning Using Python (2026) — Certificate Code: 9686515
- Data Analytics with Excel, Power BI, Python & SQL — Tutorials Point BD (2026) — Certificate ID: 26925b25c5ae5683
- Introduction to Data Engineering — IBM / Coursera (2026) — Certificate ID: WDYVTAPD8EYC
- CCNA (Cisco) & MTCNA (MikroTik) — Creative IT Institute (2025) — Certificate ID: CMTCNA-25010205
- Data Science Essentials with Python; Data Analytics Essentials — Cisco Networking Academy
- Python Essentials 2 — Python Institute; AI Fundamentals — IBM SkillsBuild