MALEHA ISRAT CHOWDHURY

St. John's, NL, Canada. • malehaickheya@gmail.com • +1 (548) 883-7254
linkedin.com/in/maleha-ick • malehaisratchowdhury.netlify.app

SUMMARY

Completed my **Master's in Computer Engineering**, gaining hands-on experience across the full ML pipeline, including feature engineering (TF-IDF, bigrams) and deployment via Django and REST APIs. Co-authored two peer-reviewed **Springer Nature** papers on fake news and deceptive review detection, using a dataset of 74,428 news articles and a separate dataset of 1,600 hotel reviews, applying models such as Logistic Regression, SVM, Bi-LSTM, and Random Forest, achieving up to 98.82% accuracy and 92.19% validation accuracy. As a **Junior Data Scientist** delivered financial crime analytics by building 5+ anomaly detection models and producing 10+ validation reports and as a Tech Assistant migrated 10,000+ records to SQL and reduced manual review time by 45% through Python automation.

EDUCATION

Master of Applied Science (MASc.) | Computer Engineering Sep 2023 – May 2025
Memorial University of Newfoundland, St. John's, NL
Bachelor of Science (BSc.) | Computer Science and Engineering Jan 2018 – Jan 2022
East West University, Dhaka, Bangladesh

PROFESSIONAL EXPERIENCE

Junior Data Scientist | R N Trading Ltd., Dhaka, Bangladesh Jul 2022 – Aug 2023

- Delivered financial crime analytics across 3+ client engagements, translating AML and trade surveillance requirements into scalable data-driven solutions.
- Designed and developed 5+ anomaly detection models to identify market manipulation behaviors such as spoofing, layering, and wash trading using transactional trading data.
- Collaborated with cross-functional stakeholders to align model outputs with compliance expectations and business objectives.
- Actively contributed as a reliable team member, balancing project demands with tight timelines while supporting collaborative problem-solving and knowledge sharing.

Tech Assistant | East West University, Dhaka, Bangladesh March 2022 – June 2022

- Migrated 10,000+ records from unstructured spreadsheets into a normalized SQL database, preserving referential integrity.
- Built Python validation scripts reducing manual data review time by 40–50% through automated error detection.
- Documented migration workflows, enabling repeatable processing and faster team onboarding.
- Worked collaboratively as part of a team, adapting quickly to challenges and maintaining efficiency under time constraints.

RESEARCH & PUBLICATIONS

Enhancing False News Detection Through ML and Transfer Learning
ICDMIS 2024, Springer Nature Oct 2024

- Processed tokenization, stopword removal, normalization, and class balancing on a big dataset of 74,428 news items.
- Developed several models (Logistic Regression, SVM, and Bi-LSTM) and tested the performance across deep learning and conventional machine learning techniques.
- Bi-LSTM demonstrated excellent performance and unexplored data sources, achieving 98.82% validation accuracy.

- Evaluated model decisions and chose architectures based on their ability to represent patterns rather than just their accuracy.

Detection of Deceptive Hotel Reviews Through Machine Learning
Springer Nature 2024 Jul 2024

- Designed and implemented an ML-based framework to classify 1,600 hotel reviews as genuine or deceptive using the Deceptive Opinion Spam Corpus dataset.
- Built a complete NLP pipeline that includes model assessment, TF-IDF feature engineering and text preparation.
- Demonstrated that Random Forest outperformed other models for opinion spam detection, achieving **92.19% accuracy.**

ACADEMIC PROJECTS

House Price Prediction | Python, scikit-learn, XGBoost, pandas MASc Capstone, 2025

- Built and compared Linear Regression, Ridge, and XGBoost models on structured real estate features to predict property sale prices.
- Reduced RMSE by 18% through data cleaning and feature engineering alone handled missing values, encoded categorical, and removed transaction outliers that were skewing the regression line.
- Used Ridge regularization to control overfitting on correlated features like square footage and neighborhood; evaluated all models using RMSE and R^2 with cross-validation.

Dengue Incidence Forecasting | Python, scikit-learn BSc Capstone, 2022

- Forecasted weekly dengue cases across two cities (San Juan, Iquitos) using 1,456 records and 24 meteorological features per city.
- Built and benchmarked KNN, Random Forest, Gradient Boosting, and SVR; evaluated using MAE before and after correlation-based feature selection
- Reduced feature count from 16 to 11 and improved KNN MAE from 19.90 to 16.88 on San Juan, outperforming published RF and GBR benchmarks on the same dataset by choosing the simpler model that fit the data structure.

PERSONAL PROJECTS

Real-Time Crypto Sentiment Analysis | Python, NLP, VADER, Hugging Face Transformers, APIs 2026

- Building an NLP pipeline that combines VADER and a fine-tuned transformer to analyze live Reddit data and track crypto market sentiment in real time.
- Integrating CoinGecko API for live price feeds; designed to surface sentiment-price divergence as an early trading signal.
- Deploying as a Streamlit dashboard for live demo and reproducibility.

TECHNICAL SKILLS

Languages: Python, SQL, C/C++, PHP, HTML/CSS
Machine Learning & AI: scikit-learn, TensorFlow, PyTorch, XGBoost, Random Forest, SVM, Bi-LSTM, Logistic Regression, Ridge Regression, KNN, Gradient Boosting
NLP: TF-IDF, VADER, Text Preprocessing & Feature Extraction
Data & Analytics: Pandas, NumPy, Feature Engineering, Data Cleaning & Processing
Databases: MySQL, PostgreSQL
Web & Deployment: Django, REST APIs, ML Pipelines, Git