

MST. ROKEYA KHATUN

Machine Learning & Data Science · Software Engineer

Dhaka, Bangladesh · +880 1793-368956 · khatunrokeya406@gmail.com · github.com/ROKEYAKHATUN · [linkedin.com/in/rokeya-sriti](https://www.linkedin.com/in/rokeya-sriti)



SUMMARY

Computer Science & Engineering graduate (CGPA 3.82, AIUB) and software engineer by training, specializing in machine learning, deep learning, and data science. Built and published six end-to-end ML projects covering NLP classification, computer vision, clustering, and healthcare prediction, with every metric measured and reproducible from public notebooks. Co-author of a few-shot learning paper submitted to DICTA 2026, and an undergraduate thesis that pushed fine-tuned GPT models to 86.7% accuracy on IMDB sentiment analysis. Strong software-engineering fundamentals (OOP, DSA, the same platform built in three stacks); chose to specialize in ML/DL and currently learning modern AI.

TECHNICAL SKILLS

ML & Data Science: Python, Scikit-learn, TensorFlow/Keras, Pandas, NumPy, Matplotlib, Seaborn, NLP, Computer Vision, Streamlit, Jupyter, Google Colab

Concepts: Feature Engineering, Cross-Validation, Model Evaluation, CNNs, Clustering, DSA, OOP

Languages: Python, Java, JavaScript, C++, C#, HTML, CSS

Web (working knowledge): React, Next.js, NestJS, Spring

Databases & Tools: MySQL, PostgreSQL, Git/GitHub

Currently learning: Modern AI: LLM-based tooling and applications

EDUCATION

BSc in Computer Science & Engineering 2020 - 2025

American International University-Bangladesh (AIUB)

CGPA 3.82 / 4.00 · Thesis: fine-tuning GPT foundation models

Higher Secondary Certificate (HSC) 2019

Cambrian School and College, Dhaka

Secondary School Certificate (SSC) 2017

Police Line's School and College, Kushtia

RESEARCH

Training-Free Whitened Local-Global Fusion for Few-Shot Plant Disease Recognition

2nd Author · Submitted to DICTA 2026, under review · Supervisor: Md. Ismail Hossen, Griffith University, Australia · [LinkedIn](#) · [Google Scholar](#)

Engineered a fully training-free framework on a frozen ResNet-18, combining ZCA-whitened local descriptors with a global Mahalanobis metric to bridge the laboratory-to-field domain gap in few-shot plant disease recognition. No fine-tuning required.

Advancing NLP Tasks Through Fine-Tuning Foundation of GPT Models

Undergraduate Thesis, AIUB · 2024

Developed adaptive few-shot learning techniques on top of GPT foundation models, improving the efficiency of the fine-tuning pipeline and reaching 86.7% accuracy on IMDB sentiment analysis with small labelled corpora.

PROJECTS

[Emotion Detection from Text](#) Python · Scikit-learn · TF-IDF · LinearSVC · GridSearchCV

Multi-class NLP pipeline that classifies text into six emotions. Benchmarked Naive Bayes, Logistic Regression, and LinearSVC on the same split, tuned the winner with GridSearchCV to 90.5% test accuracy, and persisted the model behind a real-time prediction function.

[CIFAR-10 Image Classification](#) TensorFlow / Keras · CNN · Streamlit

Convolutional neural network trained on 60,000 images across ten classes. Data augmentation and dropout lifted accuracy from 69.3% to 75.6%, and the trained model is served through a Streamlit app with upload-and-predict plus confidence scores.

[Breast Cancer Detection](#) Python · KNN · StandardScaler Pipeline

Supervised classifier for the Wisconsin diagnostic dataset (569 samples, 30 features). Systematic K tuning across 20 values inside a scaling pipeline reached 98.25% test accuracy with 100% precision on the malignant class.

[Airbnb Market Segmentation](#) Python · K-Means · Scikit-learn · Pandas

Unsupervised study of 48,895 NYC listings. K-Means (K=5 via the elbow method, validated by silhouette score) segments the market into five named clusters, translated into pricing and occupancy insights for hosts.

[Resume Classification System](#) Python · CountVectorizer · Multinomial Naive Bayes

NLP system that predicts a candidate's job category from raw resume text, trained on 13,389 resumes across 43 categories. Evaluated with accuracy (71.3%), a full classification report, confusion matrix, and 5-fold cross-validation.

[Loan Approval Prediction](#) Python · Decision Tree · Logistic Regression · KNN

Three-model comparison evaluated with k-fold cross-validation rather than a single split. Logistic Regression won on accuracy (80.5%), and feature analysis identified credit history as the strongest predictor of approval.

[Recruitment Management System \(software\)](#) Next.js + NestJS/PostgreSQL · Java Spring/MySQL · Python · [Java Spring](#) · [Python](#)

The same hiring platform built three times in three different stacks, with four user roles and one shared domain design: candidates apply, companies post jobs, recruiters interview and hire, admins manage the system.

EXPERIENCE

Intern, The Premier Bank PLC Gulshan Branch, Dhaka · Jun 2024 - Sep 2024

- Completed a three-month internship embedded in the Gulshan branch's daily operations.
- Worked on the bank's challan (payment processing) workflow alongside senior staff.
- Hands-on with the bank's official software supporting real day-to-day banking operations.

REFERENCES

Sharfuddin Mahmood

Assistant Professor & Special Assistant [OSA], CS, American International University-Bangladesh

smahmood@aiub.edu · +880 1681-250410

Syed Shafqat Rabbi

EVP, Gulshan Branch, The Premier Bank PLC

shafqat@premierbankltd.com · +880 1755-500040