

Istiaq Islam

Software Engineer · Data Engineer

Dhaka, Bangladesh · +880 1755 578878 · isttiak@gmail.com
isttiak.vercel.app · github.com/isttiak · linkedin.com/in/isttiak

SUMMARY

Computer Science graduate (BRAC University) working across full-stack software engineering and data. Built and deployed **three production web applications** (React, TypeScript, Node.js, PostgreSQL, MongoDB) and an **end-to-end machine learning pipeline** (LightGBM, NLP, FastAPI) as capstone of a government-backed data science program. Undergraduate thesis applied Bayesian statistics and sentiment analysis to e-commerce rating systems. Seeking software engineering and data engineering roles.

TECHNICAL SKILLS

Languages: Python, SQL, TypeScript, JavaScript, Bash

Data Science & ML: Pandas, NumPy, Scikit-learn, LightGBM, XGBoost, SMOTE, hypothesis testing, Bayesian methods, time series (ARIMA, Prophet), NLP (TF-IDF), Matplotlib, Seaborn, Plotly

Web & Backend: React, Next.js, Node.js, Express, FastAPI, REST APIs, Tailwind CSS, TanStack Query, Zustand

Databases: PostgreSQL, MySQL, MongoDB, Supabase, Firebase, SQL query optimization

Data Engineering & DevOps: Apache Spark, Hadoop, Airflow, Docker, Kubernetes (coursework), Git, CI/CD, Linux, Jupyter, LaTeX

PROJECTS

Mercari Price Suggestion — End-to-End ML Pipeline 2026
Python, Pandas, Scikit-learn, LightGBM, NLP, FastAPI · [GitHub](#) Capstone, ITS 507

- Built a complete regression pipeline predicting marketplace listing prices from title, description, category, brand, and condition on the Kaggle Mercari dataset (large-scale, sparse NLP features)
- Achieved **RMSLE 0.465 and MAE \$10.44** with LightGBM — a **25.5% improvement** over a Ridge regression baseline; kept the full feature matrix sparse end-to-end for laptop-scale training
- Deployed the selected model behind a **FastAPI** endpoint with an inference-time preprocessing module that exactly mirrors the training pipeline, preventing train/serve skew

FinTrack — AI-Powered Personal Finance App 2025
React 19, TypeScript, Supabase (PostgreSQL), TanStack, Groq AI · [live](#) · [GitHub](#) Live in production

- Designed and shipped an open-source finance platform covering expense tracking, lent/debt ledger, and investment portfolio, self-hostable and built for the Bangladeshi context
- Integrated **10 AI features** via the Groq API (smart categorization, anomaly detection, budget analysis, natural-language chat over user data) with compact pre-aggregated prompts for free-tier costs
- Enforced per-user data isolation with PostgreSQL **Row Level Security**; added CSV import with column-mapping preview, full Excel/CSV export, and soft-delete

Ihsan — Islamic Productivity Platform 2025
React, Node.js, Express, MongoDB, Firebase Auth · [live](#) · [GitHub](#) Live in production

- Built a full-stack habit platform (prayer, zikr, fasting, Quran tracking) with streaks, goals, and per-habit analytics dashboards; 19-test backend suite
- Engineered timezone-correct daily buckets where the client's local civil date is authoritative — accurate local-midnight resets worldwide with zero user configuration
- Optimized write-heavy counters with debounced client batching and atomic MongoDB operations (\$inc), eliminating lost updates; implemented lazy streak expiry requiring no cron infrastructure

Adaptive Rating System for E-commerce — Undergraduate Thesis

2024 — 2025

Bayesian statistics, sentiment analysis, Chrome extension · [GitHub](#)

Supervisor: Prof. Farida Chowdhury

- Designed a manipulation-resistant product rating algorithm combining **Bayesian averaging**, exponential time decay, and AI sentiment extraction from review text
- Solely responsible for all data work: EDA, hypothesis testing, algorithm design, and statistical validation; contribution commended by supervisor for originality and statistical rigor
- Shipped the algorithm as a Chrome extension that recomputes robust ratings directly on live e-commerce product pages

Brain Stroke Classification

2023

Python, Scikit-learn, SMOTE · [GitHub](#)

Team of 4, CSE422

- Preprocessed 43,400 clinical records with heavy class imbalance: missing-value imputation (14k+ nulls), categorical encoding, and **SMOTE** resampling
- Implemented logistic regression (89.7% accuracy) and compared 4 classifiers with 10-fold cross-validation; best model (Random Forest) reached **96.8% accuracy**

EDUCATION

BSc in Computer Science, BRAC University

Sep 2021 — Nov 2025

CGPA 3.30 / 4.00 · 6 credits from Bard College at Simon's Rock (international exchange)

Dhaka

- Coursework: Statistics & Probability, Linear Algebra, Numerical Methods, Database Systems, Artificial Intelligence, Algorithms, Data Structures

Higher Secondary Certificate (HSC), Notre Dame College

2018 — 2020

Science stream

Mymensingh

Secondary School Certificate (SSC), Milestone School

2016 — 2018

Science stream

Uttara, Dhaka

CERTIFICATIONS, TRAINING & LANGUAGES

Certificate on Data Science (ITS 507)

Feb 2026 — Present

BRAC University · Government of Bangladesh ICT Division · Co-funded by Asian Development Bank

Dhaka

- Admitted through competitive national written exam and interview
- Curriculum: SQL, advanced ML (XGBoost, LightGBM), Bayesian statistics, time series, big data (Hadoop, Spark), MLOps (Airflow, Docker, Kubernetes, AWS/Azure/GCP), NLP; concludes with industry internship
- LFD201: Introduction to Open Source Development, Git, and Linux — The Linux Foundation
- HTML5 and CSS Fundamentals — edX / W3C
- Languages: Bengali (native), English (professional — IELTS Academic 6.5)

AWARDS & ACTIVITIES

- **National Finalist**, imaGen Ventures Youth Challenge 2024 (UNICEF Bangladesh & JAAGO Foundation) — SDG-aligned application; led data collection and market analysis
- **Co-founder & Executive Volunteer**, ARUN non-profit (2019 — present) — 5+ years leading disaster-relief fundraising campaigns
- **Executive Member**, BRACU Robotics Club (2023 — 2024)