

SHAH MD ABIR HUSSAIN

(929) 204-9535 | Abir.h.shah@gmail.com | linkedin.com/in/abirshah | github.com/shah-abir

PROFESSIONAL SUMMARY

MS Financial Technology candidate with hands-on experience building ML-driven data pipelines and credit risk models for financial services. Skilled in Python, AWS, XGBoost, and SHAP-based explainability, with domain knowledge spanning ETL architecture, regulatory compliance (ECOA, FCRA, GDPR), and quantitative analysis. Seeking data engineering and analytics roles in investment banking and fintech.

SKILLS

Programming Languages: R, Python, MySQL, PostgreSQL, Git, Shell

Libraries & Frameworks: scikit-learn, XGBoost, SHAP, NumPy, Pandas, Matplotlib, Seaborn, SciPy, PySpark, Streamlit

Cloud & Infrastructure: AWS (S3, Glue, Lambda, EC2, Athena, Airflow, IAM, Textract)

Tools & Platforms: Power BI, Tableau, Jupyter Notebook, JMP, GitHub, Google Suite, Jira, R Studio, SAS, Bloomberg Terminal

EDUCATION

University of Connecticut — *MS Financial Technology*

May 2025

Courses: Deep Learning, Data Mining, Time Series Forecasting, Predictive Modelling, Financial Programming & Modelling.

City University of Hong Kong — *BEng Mechatronic Engineering*

Oct 2019

Certificates: Associate Data Engineer in SQL (DataCamp) | Python & Statistics for Financial Analysis (Coursera) | Python for Data Science, AI & Development (Coursera) | Multivariate Calculus for ML (Coursera)

WORK EXPERIENCE

AutoShopIQ

Stamford, CT

AI & Data Pipeline Engineer

Sept 2025 – Present

- Architected an end-to-end, event-driven data ingestion pipeline using **AWS S3, Lambda, and IAM** to automatically capture, process, and store automotive repair documents (PDFs, invoices, inspections).
- Built Document AI workflows leveraging **Amazon Textract**, OCR tooling, and Python to extract and normalize unstructured repair data into canonical JSON schemas for analytics and ML.
- Designed AI-ready data pipelines integrating validation, transformation, and **PostgreSQL** storage to support downstream machine learning, retrieval, and recommendation systems.
- Improved pipeline reliability by adding structured logging, error handling, and retry mechanisms for document ingestion and AI extraction processes in real shop environments.

Fixnox

Sydney, Australia (Remote)

Data Engineer Intern

Jan 2024 – Dec 2024

- Architected a batch ETL pipeline on **AWS (S3, Glue, Athena)** to ingest and analyze FIX protocol trade messages across multiple counterparties, following a **medallion architecture** (Bronze/Silver/Gold) for progressive data refinement.
- Developed a Python-based FIX message data generator producing ~23,000 realistic trade logs (NewOrderSingle, ExecutionReport, Rejects) partitioned in Hive-style format by date for cost-efficient querying in Athena.
- Engineered ETL transformations with **PySpark and Pandas** to parse raw CSV trade logs into optimized Parquet, computing derived trading metrics including rejection rates by counterparty, execution latency percentiles, and daily notional trends.

HKexcel Education Limited

Hong Kong

Physics & Mathematics Instructor

Nov 2019 – Jul 2023

- Delivered instruction to 70+ IB Mathematics and Physics students, resulting in consistent scores above 80%.
- Translated complex concepts into digestible modules, increasing student comprehension and achievement by 25%.

RELEVANT PROJECTS

Credit Risk Scoring Engine with Explainable AI — GitHub

- Built an end-to-end credit risk classifier using **XGBoost** on 10,000 synthetic credit records with 17 features (14 raw + 3 engineered), achieving a **ROC AUC of 0.77** on a held-out test set of 2,000 records.
- Integrated **SHAP (TreeSHAP)** explainability producing per-prediction waterfall plots, force plots, beeswarm visualizations, and plain-language risk factor summaries for regulatory transparency.
- Designed the system to address **ECOA, FCRA, GDPR Article 22, and SR 11-7** requirements by generating rank-ordered adverse action factors with quantified impact magnitudes.
- Deployed as an interactive **Streamlit** dashboard with real-time scoring across 14 borrower parameters; authored a 19-page white paper with 34 citations covering methodology, evaluation, and regulatory compliance.

Monte Carlo Portfolio Optimization — GitHub | **Predicting Problematic Internet Use** — GitHub

- Simulated buy-and-hold vs rebalancing strategies (2017–2022), improving risk-return tradeoff through monthly rebalancing.
- Built a classification model for adolescent internet risk using the HBN dataset; achieved an F1 score of 0.73.