Shlok Thakkar

+1 217-255-3034 | shlokat2@illinois.edu | linkedin.com/in/shlok-thakkar | github.com/shlok1806 | Portfolio

Education

University Of Illinois Urbana-Champaign
Bachelor of Science in Information Science + Data Science, Minor in Computer Science *GPA: 4.0*

Experience

Research Assistant Jan 2024 – Present
PI: Prof. Tatyana Deryugina, Department of Economics, UIUC *Champaign, IL*

- Developed an entity resolution pipeline to match company names between the Tender Electronics and Orbis databases, processing a combined dataset of over **10M records**
- Designed a custom string-matching algorithm using NLP techniques (Levenshtein distance, TF-IDF vectorization) to achieve **95% matching accuracy**, validated against a manually-labeled set
- Refactored the pipeline with memory-efficient data chunking and optimized vectorization, cutting runtime by **73%** (48 to 13 hrs) and enabling the entire process to run on a single machine

Data Science Intern June 2025 – August 2025
IQM Corporation *Ahmedabad, India*

- Owned the end-to-end development of an internal recommendation tool, enabling teams to find personality matches from a database of **50,000+** profiles
- Architected and deployed a Python-based similarity engine API on AWS, using vector embeddings to serve real-time matching requests
- Optimized vector search queries in Athena to achieve sub-**500ms** latency, reducing average query time by **18%** from the initial baseline
- Developed a full-stack interface using Streamlit, allowing non-technical users to interact with the engine and visualize results

Technical Lead, Student Project Jan 2025 – Present
Project Lumyn @ UIUC *Champaign, IL*

- Serve as the technical lead for a 5-member student team building a predictive maintenance system from **10K+ hours** of flight telemetry data
- Personally developed the deep learning model in **PyTorch**, architecting the data pipeline and achieving **93% accuracy** in battery failure prediction
- Built and deployed a REST API using FastAPI to serve model predictions, successfully stress-testing it to handle a load of over 1,000 requests/minute

Projects

Credit Risk Prediction System | *Python, SQL, XGBoost, SHAP, Tableau* | GitHub Dec 2024

- Trained an XGBoost model on **50K loan applications** to predict defaults, achieving **94% AUC** on the hold-out set
- The model successfully identified a high-risk group of applicants that accounted for over **75%** of the total potential charge-offs in the test set
- Utilized SHAP for model interpretability and developed a Tableau dashboard to monitor fairness metrics (e.g., Disparate Impact ratio) in alignment with Fair Lending principles

Options Pricing & Risk Management Suite | *C++* July 2025 - Present

- Implemented Black-Scholes (analytical) and Monte Carlo (simulation) pricing models in C++ for European options
- Utilized C++ multithreading to parallelize the Monte Carlo simulation, achieving over **100x speedup** over a single-threaded Python baseline
- Engineered a real-time calculator for key option Greeks (Delta, Gamma, Vega, Theta) for risk management analysis

Technical Skills

Programming: Python (NumPy, Pandas, Scikit-learn), SQL, R, Java, C++
ML/Analytics: PyTorch, XGBoost, Statistical Modeling, A/B Testing, Causal Inference
Visualization: Tableau, Power BI, Matplotlib, Plotly Dash, Streamlit
Finance/Risk: Credit Risk Modeling, Options Pricing, VaR/CVaR, Fraud Detection, Basel III, CECL