Shubham Madhavi
Software Engineer (Data & AI)
Jersey City, NJ | (201) 526-5040 | shubhammadhavi9@gmail.com | github linkedin

Professional Summary

Data-driven Software Engineer with a Master's in Computer Science and hands-on experience in quantitative finance and machine learning. Proven ability to build and optimize data-intensive systems, from ETL pipelines and backtesting frameworks to high-accuracy ML models. Seeking to apply expertise in Python, data engineering, and AI to solve complex, high-impact problems.

Technical Skills

- **Languages:** Python (Expert), SQL (Expert), Java, C++
- **Data Science & ML:** scikit-learn, TensorFlow (Keras), Pandas, NumPy, NLP (TF-IDF, Word2Vec, GloVe)
- **Data Engineering:** PySpark, Hadoop, Airflow, ETL/ELT, SQL (MySQL, PostgreSQL), NoSQL (MongoDB)
- **Cloud & DevOps:** AWS (SageMaker), Docker, Git, Linux, CI/CD, Azure (Familiar)
- **Generative AI:** RAG, LangChain/LangGraph, Hugging Face Transformers, Vector DBs (FAISS, Pinecone)
- **Web & Visualization:** Node.js, React, Matplotlib, Seaborn, Plotly, Power BI, Tableau

Education

Pace University | New York, NY *Master of Science in Computer Science* | Sep 2022 – May 2024

- **GPA:** 3.66/4.0
- **Relevant Coursework:** Artificial Intelligence, Algorithms & Computing Theory, Database Management Systems, Parallel Computing.

Mumbai University | Mumbai, India *Bachelor of Engineering in Computer Engineering* | Aug 2018 – May 2021

- **Honors:** Project Planet USA Winner ($15k grant for developing bio-bricks to restore coral reefs).

Professional Experience

Python Developer Intern | **Baker Trading** | Seattle, WA *Oct 2024 – Jun 2025*

- Improved predictive trading model accuracy by **20%** by architecting and implementing a comprehensive backtesting framework in Python.
- Engineered and maintained robust ETL pipelines (Python, Pandas) to clean, transform, and load large-scale financial market data, supporting real-time analytics and model training.
- Developed and deployed automated trading scripts leveraging financial data APIs, directly contributing to optimized trade execution strategies.
- Optimized complex SQL queries and designed efficient database schemas (PostgreSQL) to store and query high-frequency, real-time trading data.
- Collaborated with quantitative traders to build and refine predictive models using scikit-learn and TensorFlow to forecast market trends.

Software Developer | **Safeline Electricals** | Mumbai, India *Aug 2021 – Aug 2022*

- Automated key business reporting by building a centralized Python/SQL ETL warehouse, eliminating 20+ hours of manual data consolidation per week and enabling daily KPI dashboards.
- Developed SQL-based financial models and KPI dashboards (D3.js) to track project variance and forecast spend, providing managers with real-time decision-making power.
- Prototyped an anomaly detection system (Python, scikit-learn) to flag outlier spending, enabling early intervention before month-end close.

Projects

Credit Card Fraud Detection (Machine Learning) | GitHub Link

- Achieved **92% accuracy** in fraud detection by engineering a Random Forest model on a 284,000-transaction dataset.
- Boosted fraud recall by **25%** by implementing SMOTE to resolve severe class imbalance, ensuring more fraudulent transactions were caught.
- Conducted in-depth EDA to identify key fraud indicators and engineered new features based on transaction amount and time-based trends.

Restaurant Review Sentiment Analysis (Deep Learning) | GitHub Link

- Attained **92% accuracy** in sentiment classification by designing and training an LSTM-based deep learning model, outperforming traditional ML baselines (SVM, Random Forest).
- Built a complete data processing pipeline using TF-IDF and word embeddings (Word2Vec, GloVe) to convert raw text into feature vectors for nuanced analysis.