Samanvay Malapally Sudhakara

+1 217 926 6744 | smalapallysudhakara@gmail.com | LinkedIn | Project Portfolio | Github | Kaggle

Experienced Data Scientist with **3+ years** of experience leveraging advanced analytics & machine learning techniques to solve complex challenges in finance, technology, healthcare & business operations. Adept at deriving actionable insights from complex datasets using statistical analysis & predictive modeling. Skilled in various programming languages & data management tools, capable of creating effective data visualizations & communicating insights to stakeholders. Proficient in cloud technologies & experienced in deploying scalable data solutions. Familiar with SDLC methodologies such as Agile & Waterfall. Strong quantitative & mathematical skills, including statistical modeling, stochastic calculus, numerical methods, game theory & expertise in building algorithmic trading systems & developing high-frequency trading (HFT) strategies.

TECHNICAL SKILLS

Languages	Python, C++, R, SQL, NoSQL, Linux Shell, Git
Data Visualisation	Numpy, Pandas, Seaborn, Matplotlib, Plotly, ggplot2, Tableau, Power BI, Streamlit, Excel (Pivot Tables, VLookup)
Machine Learning	Regression, Classification, Decision Trees, Ensemble Trees, KNN, K-Fold CV, PCA, LDA, SVM, ANN, CNN, RNN, LSTM, GRU, Transformers, DQN, DDPG, VAE, GANs, U-NET Diffusion Models, NLP
ML Packages	Pytorch, Tensorflow, Keras, Scikit-learn, OpenCV, Transformers, Huggingface-hub, NLTK, Word2Vec, spaCy, LangChain
Cloud & Databases	AWS (S3, EMR, EC2, Redshift, Sagemaker), GCP (Compute Engine, GCS, Vertex AI, Cloud Run, BigQuery), MS Azure
Big Data, CI/CD	MySQL, Apache (Hadoop, Airflow, Spark, Kafka), MapReduce, MongoDB, Jenkins, Kubernetes, Docker, Databricks, SnowFlake
Quantitative Skills	Quantitative Finance, Stochastic Calculus, Derivatives, Hypothesis Testing, Statistical Analysis, Risk Management, Market Microstructure, Algorithmic Trading, Interest Rate Models, Game theory, Dynamic Programming & Reinforcement Learning
Certifications	Deep Learning Specialisation, Machine Learning Specialisation (*deeplearning.ai*), Bloomberg Market Concepts

PROFESSIONAL EXPERIENCE

Data Scientist, *Discover Financial Services, IL* Jan 2024 - Current

- Developed & deployed advanced credit risk assessment models using scikit-learn for tree based models & TensorFlow for deep learning models, leading to a 30% increase in the accuracy of predicting loan defaults & reducing the overall risk profile.
- Enhanced sentiment analysis for market sentiment & creditworthiness assessments using Neural Network Models achieving a 30% improvement in performance.
- Implemented real-time data processing pipelines using Apache Spark & Kafka, which reduced data latency by 25% & improved the timeliness of credit risk evaluations & decision-making.
- Streamlined data warehousing & retrieval processes by implementing AWS Redshift, resulting in a 35% increase in query performance & enabling more efficient financial risk reporting & analytics.
- Optimized SQL queries on MySQL databases, resulting in a 25% improvement in data retrieval speed for financial reporting & analysis, ensuring timely access to critical business insights.

Quantitative Analyst (Practicum, Internship), *JIA Finance, NY* Jan 2023 - Aug 2023

- Automated Mortgage Guidelines Interpretation leveraging OpenAI GPT 3.5, GPT 4 APIs, LangChain, FAISS/Chroma DB vectorstore for rapid ingestion & validation of documents resulting in enhanced speed & accuracy of mortgage processing.
- Engineered specialized task-specific Retrieval Augmented Generation (RAG) Q&A chains that refined query processing using a recursive tree approach combined with LLM-based memoization which significantly increased response accuracy for queries boosting zero shot accuracy by 30 %.
- Exploratory Data Analysis on over 100M Fannie Mae mortgage records using Tableau along with Data Cleaning & Feature Engineering with AWS Sagemaker & Redshift Connector, applying various other data manipulation & visualization tools for actionable insights.
- Modeled loan survival & default rates curves employing the Cox Proportional Hazard Model & Kaplan Meier Estimator & assessed the impact of various macroeconomic factors such as changes in unemployment rates & income on default & delinquency rates.
- Crafted a comprehensive cash flow model for Mortgage Backed Securities (MBS) integrating credit default, prepayment, & loss severity models with a yield curve calibrated Cox Ingersoll Ross (CIR) Interest Rate model.

Data Scientist, *HCL Technologies, India* Jul 2020 – Jul 2022

- Applied NLP techniques to analyze news sentiment & financial reports, extracting insights to inform investment decisions & enhance portfolio performance, & developed deep learning models (TensorFlow) for sentiment analysis & topic modeling.
- Utilized advanced statistical methods & time-series forecasting models to analyze historical market data, identify patterns, predict market trends, & provide insights to optimize trading strategies & improve portfolio performance.
- Collaborated with financial analysts & data engineers to gather, prepare, & visualize data using Power BI, Docker, & AWS, ensuring consistency & repeatability in deployment, & facilitating data-driven investment decisions.
- Utilized Docker containers to encapsulate ETL processes and dependencies, ensuring consistency and repeatability across different environments and simplifying deployment to cloud platforms such as AWS.
- Documented A/B testing methodologies, results, & insights comprehensively for regulatory compliance & future reference. Presented findings in clear, actionable reports highlighting the improvements of the models.

PROJECTS

High Frequency Trading Backtesting Pipeline with RL Trading Bot

- Constructed Data Parsers for tick, depth update & market by order data from sources such CME Globex DataSet, IEX , Binance & Oanda along with data cleaning & exploratory data analysis to identify outliers, missing data & visualise the data (link)
- Simplified running strategy backtests by building backtesting pipelines for running on command line with Linux shell scripts that seamlessly integrate various C++, python scripts necessary to run backtests, process results & tune hyperparameters (link)
- Engineered DQN & DDPG models in C++ for a market-taking agent & a market-making agent with PyTorch C++ API for the model & Sqlite for experience replay store & backtested using previously crafted pipelines (link)

Forex Ladder Trading Strategy *(practicum sponsored by BP Trading)*

- Enhanced return potential by developing Forex Trading Strategy leveraging quantitative techniques for grid & lot sizing optimisation
- Sped up backtest up to 2000x using Just-In-Time Compilation on tick data from various major currency pairs
- Optimised Proprietary Trading Models by hyperparameter tuning using various automated techniques (Optuna, PySwarm, DEAP (Genetic Algorithms)) & utilized strategy validation techniques (Walk-Forward Analysis, Monte Carlo Simulations)

Trading Strategy Development, Testing, & Deployment

- Developed a mid-frequency trading strategy engine leveraging OANDA's V20 for Forex Markets, Binance API for CryptoCurrencies & Kite Connect API for Indian Equities for data streaming, trading & & order handling
- Architected pipelines for tick data and orderbook data warehousing on BigQuery and linked with Tableau Dashboard for quick data analytics
- Enhanced strategy engine improved to handle multiple strategies, track positions, & profits seamlessly with robust safety protocols & kill switches
- Deployed on Google Cloud Platform using Cloud Compute VMs, Cloud Storage, & Secret Manager for scalability & efficiency with capabilities for running backtests & live trading
- Built GUI with Flask to stream real-time portfolio information to webpage for manual strategy engine monitoring & shut down.

Large Language Model Projects

- LLM-Driven Stock Signal Extraction using Langchain & OpenAI's GPT 3.5 & GPT 4 APIs to extract trading signals from news RSS feeds & YFinance API for company financials to conduct financial analysis, synthesizing diverse data streams into predictive models for equity price forecasting
- Built GPT-2 (124M) from scratch & trained it on FineWeb Dataset (10B tokens) using a HPC cluster of 8 Nvidia A100 GPUs with distributed training & achieved a validation set log loss of 3.068 & perplexity of 22

Equity Options Dashboard with Streamlit

- Built an option dashboard that displays real-time option data such as recent close, greeks, IV with volatility smile to enable informed options trading
- Developed an option payoff chart constructer to forecast future option values & payoffs using Dupire Local Volatility and use advanced time series and Neural Network Models to forecast Volatility

EDUCATION

MS Financial Engineering (GPA 3.77/4), *University of Illinois at Urbana-Champaign* — Champaign, Illinois — May 2024

BE Mechanical Engineering, *Ramaiah Institute of Technology* — Bangalore, India — July 2021