Shengjun (James) Guan

+1 (812) 223-6448 // james.guan@nyu.edu // linkedin.com/in/jamesguanshengjun

EDUCATION

NEW YORK UNIVERSITY   **New York, NY**
The Courant Institute of Mathematical Sciences   *09/23 - 12/24*
M.S. in Mathematics in Finance (*Cumulative GPA:* 3.9/4)
- *Coursework:* financial data science, machine learning, computational statistics, data-driven modeling, time-series analysis, scientific computing, stochastic calculus

ROSE-HULMAN INSTITUTE OF TECHNOLOGY   **Terre Haute, IN**
B.S. in Mathematics with Double Major in Data Science (*Cumulative GPA: 3.5/4*)   *09/18 - 05/23*
- *Coursework:* operation research, deep learning, machine learning, software system design, SQL, NoSQL, OOP
- *Honors/Awards:* Dean's list 9 quarters, cum laude, Henry Turner Eddy Award for Application of Mathematics (2 students awarded out of class of 2023)

EXPERIENCE

PAVUS AI   **New York, NY**
Data Analyst Intern   *01/25 - present*
- Designed and automated data visualization for product cost price, incorporating specifications and key metrics, improving analysis efficiency by 5x
- Developed automated data analysis pipelines, streamlining insight generation and reducing twice manual effort
- Applied NLP techniques to process and clean unstructured product descriptions, improving data quality
- Leveraged machine learning models to classify and categorize products, enhancing data organization and usability
- Utilized AI LLMs to generate product categories, improving classification accuracy and scalability

SIMO Capital Holding, LLC   **New York, NY**
Quantitative Researcher Intern   *11/23 - 08/24*
- Applied KL divergence to analyze financial derivative trades and quote data, measuring informational imbalances to predict underlying asset directionality
- Generated features by applying time-series models across various traded assets combined with prior belief
- Architected and deployed a high-frequency, multi-asset research and trading platform, integrating backtesting capabilities and real-time analysis to support advanced strategy development
- Optimized implied volatility calculations with data structure, achieving a 5x increase in computational speed, significantly improving the efficiency of options pricing models
- Designed and constructed universal data pipeline that connects to data providers APIs, handling large volumes of high-frequency data, segmenting by domains, and processing it to deliver real-time analysis
- Developed an API that facilitates cloud-based backtesting with state-tracking capabilities, supported by a secured backend infrastructure using customized VPN in-networking and tunneling technology

PROJECTS

Development of an LSTM-Based Statistical Arbitrage Model (Python)
- Designed and implemented an LSTM neural network to predict excess returns of Bitcoin over SPY, optimizing model configurations and incorporating a combined adaptive loss function
- Conducted backtesting on out-sample data, evaluating performance metric such as Sharpe ratio and cumulative Pnl

Machine Learning Trading Signal Development (Python)
- Used time-series modeling, KNN, random forests, PCA on SPY500 and VIX data to predict binary one-day return, achieving 56% prediction accuracy in backtesting

NoSQL Database for Trading System (Python)
- Designed and implemented database system that stores time series stock data and textual data using Mongo, Neo4j, and Influx DB NoSQL database to improve efficiency and resistance to failure
- Developed queue system using Kafka between data provider API and database system, making system more robust

**Programming Languages*:* Python, Java, R, MATLAB, SQL   **Languages*:* English (fluent) and Mandarin (native)

**Affiliations/Certifications*:* Deep Learning Specialization on Deeplearning.ai, Passed FRM Level 1,

AI for Trading on Udacity Program, Golden Level in WorldQuant Challenge (alpha research)