Jacob Zhu

New York, NY | +1 (929)626-0940 | yz9029@nyu.edu

EDUCATION

New York University **New York, NY**
Bachelor of Arts in Data Science and Mathematics Sept. 2022 – May 2026
- **GPA:** 3.65/4.0, Dean's List
- **Major Coursework:** Machine Learning, Data Management, Principles of Data Science, Causal Inference, Discrete Mathematics, Numerical Analysis, Real Analysis, Theory of Probability

PROFESSIONAL EXPERIENCE

Soochow Securities Research Center **Shanghai, China**
Financial Engineering Intern Jul. 2025 – Present

Construction and Empirical Backtesting of Fund Managers' Forward-Looking Ability Factor
- Constructed event-window-based stock pools by selecting stocks within a 1–3 month window after disclosed events, and extracted fund performance factors using analyst upgrades and major restructuring events
- Built a forward-looking metric of active management skill from portfolio weights and performance factors
- Applied half-life weighting to incorporate the past 3 reporting periods in quarters with incomplete holdings disclosure
- Backtested decile-sorted portfolios on equity-biased hybrid funds, designed a Top-50 strategy, and demonstrated consistent outperformance over the long-term benchmark

ETF Index Enhancement Strategy (2019–2025)
- Applied hierarchical clustering with Pearson correlation to select 9 low-correlation, broad-based ETFs
- Constructed an equal-weighted initial portfolio and implemented rolling optimization (20-day window, 5-day rebalancing) to minimize squared tracking error relative to the target index
- Estimated index returns unbiasedly, with skewed, fat-tailed distribution signaling robust performance

China Construction Bank Jiangsu Branch **Nanjing, China**
Data Management Intern Dec. 2024 – Jan. 2025
- Contributed to big data risk control solutions, focusing on anomaly detection and fraud prevention algorithms (Isolation Forest, One-Class SVM) and real-time data processing (Apache Kafka, Flink)
- Participated in the 4th CCB AI Competition; used models like CatBoost, LightGBM, and AdaBoost Classifier to predict customer default risk based on 24,000 customer profiles and 6 months of credit card repayment data, achieving an AUC of 0.99 and an F1 score of 0.95 with the AdaBoost Classifier
- Conducted feature extraction and built a multi-class classification model to analyze fund investment preferences of approximately 500,000 potential clients. Optimized model performance with LightGBM using down-sampled data, achieving an F1 score of 0.64 and accuracy of 0.76

PROJECTS

2024 US Presidential Election Prediction with Machine Learning Aug. 2024 – Oct. 2024
- Analyzed historical election and polling data to predict the 2024 US Presidential Election outcome
- Explored various machine learning models such as Linear Regression, Random Forest, XGBoost, Gradient Boosting, and employed a Stacking Regressor for refined predictions; Validated predictions against historical election results

Spam Classification Using Machine Learning and Deep Learning Jul. 2024 – Aug. 2024
- Processed a dataset of spam and ham messages using tokenization, lemmatization, and stopword removal, and visualized word distributions through word clouds and bar charts
- Developed machine learning models, including Naive Bayes, Decision Tree, Logistic Regression, and Gradient Boosting, alongside a Bidirectional LSTM deep learning model for classification

Diabetes Dataset Analysis and Regression Modeling May. 2024
- Performed EDA on diabetes dataset using heatmaps, pair plots, and missing-value charts
- Developed and evaluated models including Linear Regression, Ridge, Lasso, Random Forest, and Gradient Boosting

ACTIVITIES

Peer Tutor in the Data Structures Course **Sept. 2025 – Present**
- Mentored 10+ students in understanding core data structures and algorithms concepts, including recursion, abstract data types, and complexity analysis, with applications in Java
- Led weekly tutoring sessions (Zoom and in-person) to review challenging topics and facilitate hands-on coding practice

SKILLS

Programming Languages: Java, Python
Libraries and Tools: NumPy, Matplotlib, Scikit-learn, LaTeX
Languages: English (Fluent), Chinese (Native)