Zhang Yifei

Gulou Campus, Nanjing University, No.22 Hankou Road, Gulou District, Nanjing, Jiangsu Province, China |
(+86)182-5123-9810 | hoderzyf@gmail.com | Postcode: 210093 | HomePage: hoder-zyf.github.io

Education

Department of Computer Science and Technology — Nanjing University
B.Sc. in Computer Science and Technology (FinTech) — *Sep 2020 – Jun 2024*

School of Management & Engineering — Nanjing University
M.Sc. in Financial Engineering and FinTech — *Sep 2024 – Jun 2027*

INTERN EXPERIENCE

Nanjing Securities | **Fintech Dept.** | *Software Development Engineer* — June 2022 – July 2022

- Selected key indicators from raw data tables, such as securities balance tables and customer information tables, and utilized Oracle databases for calculations on **25 indicators** across four categories (position information, investment performance, behavioral biases, and customer behavior characteristics), totaling **over 200 million records**
- Cleaned data using z-score and three standard deviation methods, applied PCA algorithm for data dimensionality reduction and used K-means clustering along with elbow method to optimize parameters
- Combined collaborative filtering and Word2vec algorithms to provide personalized security recommendations based on clients' existing holdings. Ensured recommendation accuracy and optimized recommendation strategies through information tracking and real-time feedback, thereby improving customer satisfaction and investment returns

Deloitte | **FengYu Intelligent Technology** | *Financial Report Smart Evaluation Intern* — Oct 2022 – Dec 2022

- Employed tools such as Python and Excel to meticulously clean and analyze data tables from Deloitte's extensive data pool, ensuring data accuracy, completeness, and consistency. Simultaneously, data processing procedures were continuously optimized to guarantee timely data updates, thereby enhancing data quality and real-time relevance
- Skillfully utilized Python to refine the previously manual Excel-based audit firm evaluation model, **transforming it into a streamlined automated coding implementation**, minimizing potential calculation errors caused by manual operations, and significantly boosting result accuracy and overall work efficiency
- Innovatively designed and implemented the **first-ever Python-based financial report forecasting model** grounded in financial report forecasting principles, elevating forecasting accuracy and efficiency, and supplying clients with more insightful financial data interpretation. Furthermore, a comprehensive user manual was written to aid team members in understanding and applying the model effectively

Soochow Securities | *Intern in Retail, Social Services, and Beauty Care Group* — Oct 2023 – Jan 2024

- Compiled and organized detailed information related to the **Double Eleven and Double Twelve shopping festivals**, ensuring comprehensive coverage of promotions, sales strategies, and market impact; tracked and analyzed the content and effectiveness of the "**All Girls' Offer 3**" program to identify key trends and consumer engagement
- Contributed to the creation of comprehensive in-depth reports on **Giant Biogene** and **Gambol Pet**, performing thorough research and analysis on the companies' financial performance, market positioning, and competitive advantages, which supported the final report's findings and recommendations
- Analyzed **Euromonitor databases** to extract and interpret data on the global and key country cosmetics markets, including market size, growth trends, and competitive landscape, providing valuable insights to support the company's strategic initiatives in the beauty care sector

Research & Projects

IFLYTEK Spark Camp — Aug 2023 – Aug 2023
Research and Development on Legal Pleading Generation Based on Spark 2.0 Model — *Project Member*

- Conducted comprehensive research comparing **GPT4** and **Spark 2.0**, focusing on their **architectural differences, capabilities, and application potentials** in various fields, highlighting the advantages of Spark 2.0 in understanding and generating contextually relevant content
- Developed a legal pleading generation system based on the Spark 2.0 model, designed to automate the creation of legal documents. The system integrates **advanced natural language processing techniques** to analyze case details and generate coherent, legally sound pleadings

- Implemented **customized training and fine-tuning** of the Spark 2.0 model with a curated dataset of legal documents to enhance its accuracy and relevance in legal context generation
- Evaluated the system's performance through rigorous testing with real-world legal scenarios, demonstrating a significant reduction in document preparation time and improvement in the **quality and precision** of legal pleadings

National Undergraduate Innovation Training Program Dec 2022 – Dec 2023

Volatility Prediction by Introducing Investor Sentiment *Project Member*

- Employed a diversified strategy for constructing prediction factors by integrating multiple indicators such as the phenomenon of mutual fund clustering, **the concept of Clique for institutional clustering**, **derived LSV indicators**, implied volatility indicators, and Baidu search index. This comprehensive approach aimed to accurately reflect investor sentiment and improve the accuracy and effectiveness of predictions
- For model selection and construction, we chose a **hybrid model** based on autoencoder and LSTM after evaluating the pros and cons of various machine learning models. This model effectively handles time series data and captures dynamic changes in investor sentiment. During the model construction process, we addressed numerous detailed issues such as parameter selection, handling data imbalance, and strategies for model training and testing. We successfully built a composite **LSTM-autoencoder model** and applied it to real data prediction tasks

Undergraduate Thesis (Outstanding) Dec 2023 – May 2024

Consistency or Discrepancy between Words and Actions based on Xueqiu Platform *Author*

- Leveraged empirical data from the Xueqiu platform, a prominent social trading platform, integrating NLP techniques, sentiment dictionaries, BGE-M3 word embedding models, and GPT-4 to scrutinize the alignment between investors' discourse on social media and their actual trading activities.
- Conducted a comprehensive analysis revealing a pronounced congruence between the narratives shared by investors on Xueqiu and their market behaviors. By incorporating variables such as gender and the clarity of information, elucidated the factors that drive consistency between verbal expressions and trading actions.
- Quantified the market effects of this consistency by calculating Cumulative Abnormal Returns (***CAR***) and Investment returns (***Inv_ret***), demonstrating the substantive role of social media information in financial markets.

FinLLaVA-8B Jul 2024 – Aug 2024

Open-FinLLMs: Open Multimodal Large Language Models for Financial Applications *Project Member*

- Spearheaded the development of a **multimodal extension** for large language models (LLMs) tailored to financial applications, pioneering the seamless integration of textual, visual, and numerical data to enhance the robustness and accuracy of financial modeling.
- Designed and meticulously curated a comprehensive multimodal financial benchmark dataset, focusing on **chart and table-centric financial tasks**, encompassing **800** meticulously crafted Q&A samples.
- Led the execution of an advanced multimodal instruction fine-tuning and alignment process, significantly enhancing the model' s capability to interpret and synergize text, visual elements (such as charts and tables), and numeric data, thereby boosting predictive performance in financial analysis.
- Released **FinLLaVA-8B**, achieving an impressive MMMU (Overall) score of **36.3** and an MMMU (Business) score of **30.7**.

Honors & Awards

Province First Prize, National College Students English Vocabulary Ability Competition	2021.05
Third Prize, National Academic English Vocabulary Competition	2021.06
Award of Excellence, The 12th Jiangsu Provincial College Students Knowledge Competition	2021.05
People's Scholarship of Nanjing University	2021.10
Elite Scholarship of Nanjing University	2021.11
People's Scholarship of Nanjing University	2022.11
Third Prize, 2022 National Student Data Analysis Competition	2022.12
Excellence Scholarship of Nanjing University	2023.12
Second Prize, Third "Xueshi Cup" Academic Paper Competition of Nanjing University	2024.05