Haojie Chu

+86 13700848819 | hjchuyu5@gmail.com | linkedin.com/in/haojie-chu-6b6677281

EDUCATION

University Malaya | Kuala Lumpur, Malaysia | Apr 2024

- Master of Data Science | GPA: 3.9/4.0 (with distinction)
- Coursework: Data Mining, Data Analytics, Machine Learning, Big Data Analytics, Network and Security

Zhengzhou University | Zhengzhou, China | Aug 2021

- Bachelor of Economic Statistics | GPA: 3.62/4.0 (Average Grade: 87.66/100) Rank: 6/65
- Coursework: Multivariate Statistical Analysis, Nonparametric Statistics, Economic Statistics, Time Series Analysis
- Awards: First-class scholarship, Merit student

INTERNSHIP EXPERIENCE

Shenzhen Hongxin Commercial Factoring Co., Ltd. | Data Analyst Intern | Jan 2020 - Mar 2020

- Utilized SQL and Python to collect, organize and clean multi-source data (300,000+ records) such as customer basic information, previous payment history and credit records.
- Participated in developing a credit scoring model for accept/reject decisions based on LightGBM, achieving a accuracy of 0.75 in predicting the probability of default.
- Created dynamic dashboard with Tableau to help non-technical colleagues with their business decision-making.

RESEARCH EXPERIENCE

Machine Learning-Based Stock Price Prediction and Portfolio Optimization | Jul 2023 - Jan 2024

- Conducted extensive literature review on machine learning models for stock price prediction.
- Extracted 1,200,000+ records of historical stock data (2008-2023) of S&P 500 index's constituents through web scraping with python (selenium & beautiful soup libraries).
- Performed data preprocessing including data cleansing, feature generation (10 new features), min-max normalization and label generation.
- Developed and compared multiple machine learning models (Logistic Regression, SVM, Random Forest, LSTM, CNN, CNN-LSTM), highlighting CNN-LSTM's superior accuracy of 0.71 and f-beta score of 0.64.
- Leveraged the CNN-LSTM model's stock price prediction results to formulate 3 investment portfolio strategies and carried back testing analysis to demonstrate their effectiveness in real-world investment scenarios.

Cloud-based Analytics for International Trade | Oct 2023 - Jan 2024

- Proposed a seamless cloud-based data analytics framework for U.S trade dataset (50,000+ records), including data ingestion, storage, processing, analytics, and visualization.
- Developed a data pipeline on Google Cloud Platform (GCP) using Composer, Directed Acyclic Graph (DAG), including ingesting data from sources, processing data, conducting data schema creation to BigQuery, and finally storing the results in data warehouse.
- Used BigQuery to analyze international trade patterns and built a dashboard using Tableau for visualization.

Flat Sales Price Prediction in Singapore | Mar 2023 - Jul 2024

- Utilized SAS to perform exploratory data analysis (including univariate, bivariate, and multivariate analysis) to understand the characteristics and relationships of 11 related features, and conducted data cleansing.
- Built both regression and classification models (including linear regression, logistic regression, decision tree, gradient boosting model and neural network) for prediction, achieving the best accuracy of 0.89.

SKILLS

- **Programming Languages:** Python, R, SQL
- **Frameworks & Tools:** NumPy, Pandas, Matplotlib, Sklearn, Tensorflow, Anaconda, SAS, SPSS, Eviews, MySQL, Spark, Hadoop, GCP, Tableau, Microsoft Office
- **Skills:** Statistical Analysis, Data Mining and Machine Learning, Econometrics
- **Languages:** English (IELTS 7), Mandarin (Native)