

ZHELU (Jerry) MAI

Denton, Texas | 940-595-9706 | zhelumai@my.unt.edu | [Personal Website](#) | [GitHub](#)

Education

University of North Texas

Master of Science, Business Analytics

Denton, TX

Jan 2021 – Dec 2022

Chu Hai College of Higher Education

Bachelor of Business Administration, Finance

Hong Kong, China

Sep 2012 – Dec 2016

Project Experience

Stock2Vec Research Project ([Poster](#))

Summer 2022

- Based on the framework of the previous Stock2Vec paper, built a **fine-tuned** stock embedding via **Word2Vec** and **FastText** model that can optimize predictions of companies' characteristics
- In the same prediction task of the previous paper, the fine-tuned models achieved at least a 10% improvement in performance
- The fine-tuned models can also adapt to general tasks, especially in predicting the market capitalization of the companies, leading to at most a 50% improvement

The Cricket Strike Rate Analysis Project

Spring 2022

- Converted files in HDFS to the database using **Sqoop** for creating analytics tables using **Hive**
- Used **Scala** to visualize the data and develop analytics reports in **Spark**
- Predicted the strike rate using a linear regression on numerical features, with **MAE** as the evaluation metrics of the feature importance

University Information Retrieval Chatbot

Spring 2022

- Scraped FAQs on the university websites with **BeautifulSoup**
- Cleaned and parsed conversations using **NLTK** and **spaCy**
- Extracted textual feature via **GloVe** word embedding and **Seq2Seq Bi-LSTM** using **TensorFlow**

Housing Prices Competition in Kaggle

Summer 2021

- Gained a 5% increase in accuracy by removing the outliers based on a well-developed **EDA** via **Matplotlib**
- Merged similar features for adjusting the skewed distribution that gained 2.7% improvement in performance
- Ranked Top 5% in the leaderboard with the tuned **gradient boosting** model

Work Experience

Operation Analyst

Nov 2017 – Jul 2019

Dollar General Global Sourcing Limited

Hong Kong

- Improved productivity by 25% through developing **VBA** automatic programs connecting to the **Access**, generated 5 daily reports and 7 weekly reports.
- Fetched the data from the **cloud storage** of 1010DATA for analytic reporting

Skillset

Programming Languages: Python, VBA, R, HTML5, JavaScript, Scala

Big Data: Hadoop, Spark, Hive

Database: MySQL, SQL Server, MongoDB, Access

Data Visualization: Tableau, Seaborn, Matplotlib, RStudio, QlikView

Tools: scikit-learn, Pandas, spaCy, git, PyCharm, TensorFlow, Linux