Hao ZHONG

(+86)137-9092-5831 22s010075@stu.hit.edu.cn

Education

Harbin Institute of Technology **Sep. 2022 – Jun. 2024**
M.S. in Big Data Science and Engineering | Graduate Academic Scholarship(First Class) *Harbin, China*

Dalian University of Technology **Sep. 2018 – Jun. 2022**
B.S. in Electronic Commerce | National Scholarship | Outstanding Graduate of Liaoning(1/74) *Dalian, China*

Publications & Patents

- Li T, Wang X, **Zhong H.** Cohesive clustering algorithm based on high-dimensional generalized Fermat points[J]. Information Sciences, 2022, 613: 904-931.
- Xiong X, **Zhong H**, Zhu Z, Lin H, Zhu J. A traffic thermodynamic diagram generation method and device, equipment, a storage medium and a program product: CN202010594681.4[P]. CN111966767A[2023-11-06].

Academic Experience

Cohesive Clustering with High-Dimensional Fermat Points **Jan. 2021 – Sep. 2022**
Journals of CCF-B | Programming | Experiment | Complexity Analysis | Manuscript Composition *Dalian, China*

- Enhanced the clustering algorithm by leveraging F_d*-Point* as cluster centers, determined through a **heuristic algorithm** tailored to identify optimal cluster centroids.
- Swiftly integrated the conceptual framework into the project and refined the algorithmic strategy based on empirical observations.

Credit Default Risk Identification Based on Ensemble Learning Algorithm **Nov. 2021 – Jun. 2022**
Dalian University of Technology 2022 Outstanding Undergraduate Thesis (Score: 96/100) *Dalian, China*

- **Feature Engineering:** Conducted data preprocessing, EDA, feature selection, and feature encoding.
- **Sample Balancing:** Addressed imbalanced samples using RUS, ROS, SMOTE-ENN, and Easy-Ensemble techniques.
- **Parameter Optimization:** Employed Optuna to optimize parameters for Random Forest and XGBoost, implemented various sampling methods, and stored result data using SQLite.
- **Model Interpretation:** Utilized SHAP to analyze the contribution of different features to the model output, conducting single-sample analysis, single-feature analysis, and interaction feature analysis.

Optimization Method for Neural Networks Based on K-Means. **Jun. 2020 – Apr. 2021**
10,000 RMB Sponsored by the Ministry of Education *Dalian, China*

- Constructing a **distributed Neural Network**. Leveraging clustering outcomes, multiple neural network models were formulated using Tensorflow to enhance the accuracy of regression predictions.

Professional Experience

Century Frontier Asset Management (AUM exceeding 20+ billion RMB) **May 2023 – Nov. 2023**
Data Scientist Internship *Shenzhen, China*

- **Research:** Addressed concept drift in stock volume and price data through experimentation with two distinct methods (Data Distribution Generation and Periodically Rolling Retrain) to augment the performance of existing models.
- **Engineering:** Overhauled the data development framework to expedite data development, ensuring user-friendliness and robustness. Additionally, introduced parallel computing to enhance daily data production efficiency.

DiDi Global Inc. (HXZ Dache, Taxi Business) **Jan. 2023 – Apr. 2023**
Algorithm Engineer Internship *Beijing, China*

- **Research:** Applied a Multi-Head Structure (a form of Multi-Task Learning) to consolidate five XGBoost models, aiming to conserve training model resources and streamline management.
- **Engineering:** Implemented parallel computing for price elasticity across cities and integrated the results into the operational research model solution.

DiDi Global Inc. (Voyager, Autonomous Driving Business) **Jan. 2022 – Dec. 2022**
Data Engineer Internship *Beijing, China*

- **Engineering:** Conceived and implemented a visualization infomation system, encompassing requirement analysis, selection of technical solutions, data processing, daily production, and the creation of a traffic heat map.

Others

Programming Proficiency: Python, Tensorflow, PyTorch, Shell, Spark, SQL, DolphinDB
Relevant Coursework: Machine Learning, Financial Mathematics, Time Series Analysis, Data Structures