LEI, WANG

E-mail: wanglei123@sjtu.edu.cn **Personal website:** http://wanglei123sjtu.github.io

EDUCATION

SHANGHAI JIAO TONG UNIVERSITY ***(Bachelor)*** *Aug. 2021 - Jun. 2025 (expected)*

Major: Industrial Engineering (Core GPA: 3.99/4.30, Ranking: 1/55) **Minor:** Finance

Core Courses: Stochastic Models (100), Machine Learning (99), Linear Algebra (98), Engineering Statistics (95), Operations Research (92)

Awards: National Scholarship, U.S. College Mathematical Contest in Modeling Finalist Award, Merit Student of Shanghai Jiao Tong University

SELECTED RESEARCH EXPERIENCE

Research on Equitable Hydrogen Technology Community Engagement *(Summer Research Program)* *May 2024 - Present*

Analyze major concerns of different communities regarding hydrogen fuel based on discussions obtained from social media sites such as Reddit.

Advisor: Zikun Ye, Assistant Professor of Marketing, University of Washington in Seattle

- **Data Acquisition:** Utilized Google's filter rules in combination with Reddit api to define various keywords, obtained a sufficient number of hydrogen fuel safety-related Reddit comments, and used LLM to perform sentiment labeling on the comments.
- **Basic Analysis:** Employed LDA method for topic modeling and information extraction on the comments, and conducted clustering analysis after text-embedding the comments using GPT.

Sentiment Explorational Analysis Based on DOUBAN Movie Reviews *Apr. 2024 - Jun. 2024*

Analyze Chinese movie reviews using machine learning algorithms.

- **Sentiment Classification:** Explored various methods to construct sentiment analysis models, such as relatively traditional dictionary-based tokenization methods and more advanced word embedding techniques, with Baidu's open-source Senta model as the baseline.
- **Further Exploration:** Took the Potential Film Score Manipulation into consideration, using t-tests and K-means clustering to detect self-boosting activity. Developed a movie search engine that allows users to search for open-ended paragraph statements.

Autonomous Driving Algorithm Development *Mar. 2024 - May 2024*

Fine-tune the ResNet-18 to achieve lane detection, integrating, lidar-based obstacle avoidance, and traffic light recognition.

- **Model Construction:** Collected and labeled lane-line image data from a specific track, fine-tuned a pre-trained ResNet-18 model, and incorporated lidar-based obstacle avoidance and traffic light recognition features.
- **Results Presentation:** Achieved first place in the class racing competition through continuous improvement.

Wordle Results Prediction and Classification Based on NLP *(Thesis: Joy of Wordle: Based on Arima and WordleRT)* *Feb. 2023*

- **Prediction Task:** Employed the GESD algorithm for data outlier handling and constructed a dual ARIMA model for forecasting.
- **NLP Task:** Utilized high-dimensional word vectors processed by BERT, passing through a fully connected network to KL divergence analysis, obtaining prediction results, calculating mean loss for prediction confidence, and iteratively adjusting network structures for further analysis.
- **Challenges Addressed:** Implemented the GESD algorithm found in the literature to detect and handle outliers. To mitigate the problem of a small sample size provided in the dataset, incorporated the MC Dropout Layer at the model output to reduce overfitting.

Empirical Study on the Relationship between Female CEOs and Corporate Innovation Levels *May 2023 - Jun. 2023*

Construct a fixed effects model to analyze panel data and provide an economic analysis.

- **Literature Review:** Proposed two null hypotheses *(H1: Female CEOs positively impact corporate innovation; H2: The likelihood of innovation driven by female CEOs in state-owned enterprises is lower than that in non-state-owned enterprises)* and determined the control variables such as board size, R&D, and revenue with reference to literature.
- **Model Establishment:** Utilized public data from CSMAR, preprocessed the data, and established a fixed-effects model for panel data analysis. Conducted robustness analysis, discussed endogeneity issues (*e.g.*, selection bias and individual heterogeneity), and used multi-period lagged explanatory variables to partially address the aforementioned concerns.

LEADERSHIP & ACTIVITIES

A+ Club of ME *Core Member & Academic Counselor* *Jun. 2023 - Present*

Ernst & Young (China) Advisory Limited *Intern, Corporate Tax Department* *Aug. 2023 - Sep. 2023*

SKILLS

Computer Skills: Python, LaTeX, Stata, Minitab, Excel, PowerPoint, Word

Language: Fluent in English (CET4: 623/710; CET6: 615/710) and Mandarin (Native)