YUE WU

✉ 12112422@mail.sustech.edu.cn · 📞 (+86) 153-3360-6607 · Personal Website:yuewu0301.github.io

EDUCATION

Southern University of Science and Technology, Shenzhen, Guangdong   09/2021 – 07/2025
Bachelor of Statistics,GPA: 3.80/4.00, Rank: 4/35

- Main Courses:Statistical Learning(98),Time Series Analysis(92),Operations Research and Optimization(90), Probability Theory and Mathematical Statistics(97),Statistical Computation and Software(91)

RESEARCH&PROJECT EXPERIENCE

Application of Deep Learning in Multi-Agent Financial Simulation   **Research Assistance**
Supervised: Prof. Peng Yang, Southern University of Science and Technology   *07/2023-05/2024*

- **Model Building:** Developed a **transformer-based autoencoder** with a novel calibration method, improving multi-agent LOB financial simulator's performance by over **50%**, and conducted comprehensive experiments with visualizations and **ablation studies** for model analysis.
- **Co-authored** a paper and submitted it to **NeurIPS**. **Preprint Link Here**.

Employee Performance Evaluation Using NLP and LLM   **Research Assistance**
Supervised: Prof. Wenqi Fan, The Hongkong Polytechnic University   *07/2024-present*

- **Data Augmentation:** Enhanced employee interview data using traditional NLP (random swapping, deletion, insertion, etc.) and a custom **evolution prompt-based LLM** method, achieving **1000x** data augmentation and iteratively screening by complexity and diversity.
- **Quantitative Scoring:** Developed a **few-shot** and **pair-comparing** scoring pipeline with LLMs to analyze and quantify employee interview texts for performance and promotion potential evaluation.

Personalized Knowledge Graph Summarization with LLM   **Research Assistance**
Supervised: Prof. Yujun Yan, Dartmouth   *07/2024-present*

- **Target:** Designed a **personalized knowledge graph summarization** algorithm using LLM based on historical data.
- **Contribution:** Conducted research on **Knowledge Graph Reasoning** algorithms and Question-Answering databases. Utilized all-MiniLM-L6-v2 to construct a text-based **question similarity graph** and historical data. Reproduced GreaseLM and designed baseline experiments based on this model.

Analysis of Bilibili Content Creators   **Course Project(Best)**
Supervised: Prof. Zeng Li, Southern University of Science and Technology   *11/2023-12/2023*

- **EDA:** Applied **ANOVA, two-way ANOVA, and ANCOVA** to assess the impact of tags, gender, video length, and their interactions on follower count. Utilized **GMM fitting** to classify video lengths on Bilibili and explored their relationship with follower count.
- **Feature Selection:** Performed variable selection using RandomForest, Lasso, XGBoost, and DCSIS to identify critical factors influencing follower count.
- **Regression&Clustering:** Conducted regression analysis with **weighted least squares, LightGBM, and five other models**; provided model interpretations. Employed **HDBSCAN** for clustering content creators and developed an **underestimation mechanism** to identify promising categories and video formats, offering recommendations for content creator positioning.

INTERNSHIP EXPERIENCE

LLM Training Data Quality Enhancement and Prompt Engineer   **Intern Team Leader**
Supervised: Yipeng Cheng, xDAN AI   *03/2024-06/2024*

- **Data Augmentation:** Expanded over 10 renowned datasets through data augmentation and diversification, guided by **WizardLM** and **DEITA** and other papers, utilizing advanced tools such as **Distilable** and

DataDreamer. Quantitatively evaluated LLM data based on complexity and quality using **UltraFeedBack**.
- **Prompt Engineer:** Developed diverse prompts using **CoT, Few-shot In-Context** and other methods to enhance question variety and complete high-quality specific tasks with several LLM models.

Alzheimer's Disease Image & Audio Analysis   **Research Assistance**

Supervised: Prof. Fan, Peking University Shenzhen Hospital   *09/2023-present*

- **Data Processing and Feature Engineering:** Utilized **Freesurfer** for image data processing and feature importance analysis via **RandomForest, DCSIS, XGBoost, Lasso, etc**.Utilized **OpenAI Whisper** for audio speech-to-text conversion and added manual features for subsequent LLM analysis.
- **Model Application:** Applied **LightGBM** and other 4 models for binary classification on CU-MCI patients, with hyperparameter tuning via **Bayesian Optimization**, and an **ensemble strategy**, improving performance by over **16%**. Used **HDBSCAN** for clustering, feature identification, and analysis. Innovatively constructed textual information with traditional features such as pauses and analyzed it using LLM.

EXTRA-CURRICULAR ACTIVITIES

Organizing and Participating in a Self-Supervised Learning Seminar   6/2023-8/2023

- Based on A Cookbook of Self-Supervised Learning , we **organized a seminar** to delve into the **self-supervised learning** and **meta-learning**. I studied relevant papers and conducted multiple knowledge-sharing sessions.
- Seminar Link

Organizing and Participating in a Graph Neural Networks Seminar   3/2024-4/2024

- I **organized a seminar** focused on Graph Learning, based on SC224W and Graph Representation Learning; we extensively studied **GNN, Knowledge Graph, Graph Reasoning, Graph Recommendation** and its applications, read several papers, and learned **Pytorch_Geometric**.

Chairperson of the Student Union of Zhiren (Highest-level Position)   9/2023-Present

- Collaborated with multiple student organizations to organize dozens of campus-level events and received the **First Prize** for Outstanding Student Organization (**Highest Rating**) during the tenure.
- Participated in the filming and design of the college documentary.

Captain of the University Orienteering Team and the Orienteering Club   9/2022-Present

- Organized multiple campus-level events to promote orienteering and the club received a Three-Star Club (**Highest Rating**).
- I won **third place** in the 2023 Orienteering and **fourth place** in the 2022 Orienteering at the Guangdong University Games.

AWARD

- First Class (First Place) of the Merit Student Scholarship   2023
- iGEM International Competition Gold Medal   2023
- First Prize in the China Undergraduate Mathematical Contest in Modelling in Guangdong Province   2023
- Second Price in The Chinese Mathematics Competitions in Guangdong Province   2021
- Student Cadre Model of Excellence(Top 8 at the university level)   2023
- Outstanding College Student Backbone at the University Level   2022

SKILLS & LANGUAGES

Computer Skills: Python,Java,R,SAS,Hadoop,Spark,Pytorch,Linux,Latex
Language: Native in Chinese,Proficient in English(TOFEL 103).