Hongyu Li

email: cli776@aucklanduni.ac.nz (university email), cli776776@gmail.com(personal)
Gender: Female · mobile: (+64) 0225487794

Summary

Highly motivated and results-oriented Data Scientist with a strong academic background in Computer Science and extensive hands-on experience in advanced analytics, machine learning, and data analysis. Proven track record in developing and deploying predictive models, performing data wrangling, and conducting comprehensive data analysis. Excellent communication and collaboration skills, with experience in both teaching and industry environments. Adept at leveraging cutting-edge technologies to drive business value and support strategic initiatives.

EDUCATION

University of Auckland, Master of Science (Research), *Master* 2023.03 - 2024.02

- **Natural language processing**
- **Machine Learning**
- **Code Generation**

University of Auckland, Bachelor of Science (Honours), *BSc(Hons)* 2022.03 - 2022.11

- **Computer Science**

University of Auckland, Bachelor of Science, *BSc(Hons)* 2019.03 - 2021.11

- **Computer Science**
- **Information Technology and Management**

Work Experience

Graduate Teaching Assistant, University of Auckland 2023.2 - 2024.7

Act as a Graduate Teaching Assistant for two courses:

- **Fundamentals of Database Systems**
- **Formal Specification and Design**

Key Contributions:

- Lab Supervision & Student Support: Managed lab sessions and provided real-time assistance to students, clarifying queries and resolving doubts both in and out of the classroom.
- Homework Assessment: Graded student assignments and provided constructive feedback, ensuring a fair and transparent evaluation process.
- Lecture & Examination Support to Faculty: Aided professors by preparing supplementary lecture materials and assisting in the exam proctoring process to maintain academic integrity.

Software Engineer Intern, Changshu New Changkelong E-commerce Company 2021.10 - 2022.4

- Daily maintenance: Ensured seamless operation of large advertising screens by updating content regularly, including special offers and public service announcements, to align with dynamic marketing strategies.
- UI Redesign & Feature Development: Collaborated with the team leader in revamping the user interface for the Chinese Spring Festival, ensuring enhanced user experience.
- Module Implementation: Developed and deployed new features/modules based on project requirements, contributing to the app's overall functionality.

Research

ChatGPT: Automated Code Generation: Implications for Software Development Efficiency and Accuracy 2023.2 - 2024.2

- Conducted a comparative analysis of different machine learning algorithms for code generation through a literature review.
- Evaluated the effectiveness of fine-tuning methods for large language models (LLM) in improving code synthesis accuracy and efficiency.

- Introduced novel fine-tuning approaches, including traditional team simulation and few-shot learning, enhancing code generation accuracy to approximately 73%.
- Proposed an Efficiency Quotient (EQ) for model comparison and discusses integrating diverse models within the Code Writer tool to accommodate users with varying professional backgrounds.

Data Analysis: Predictive Analysis of Stroke Patients 2022.7 - 2022.11

- Conducted predictive analysis of stroke patients using **IBM SPSS Modeller**, leveraging data mining technologies to extract valuable insights.
- Harnessing a versatile **Opensource Software Stack** encompassing Python (scikit-learn, PySpark, Spark, etc.), Tableau, Weka, AWS, and GitHub to ensure robust and holistic data analysis.
- Developed and fine-tuned predictive models to identify risk factors and potential indicators for stroke occurrence, contributing to proactive healthcare measures.

Data Analysis: Classification and Prediction of Dementia 2022.2 - 2022.11

- Performed comprehensive data preprocessing, including data cleaning, normalization, and feature engineering, to enhance the accuracy and robustness of the models.
- Implemented cross-validation techniques to ensure the reliability and generalizability of the predictive models.
- Built the machine learning models by using Python and optimised them to find the most suitable model for the dementia dataset.

Project

NLP Project: Advanced Text Sentiment Analysis 2022

- Leveraged **Meta Learning** techniques to optimize the selection of sentiment analysis methods, enhancing model adaptability across diverse data sets.
- Executed comprehensive data **preprocessing**, followed by effective feature extraction to improve model accuracy.
- Designed, trained, and **fine-tuned** a sentiment analysis model.

Data Wrangling: Relationship Between Country's Economic Status and People's Time Use 2022

- Explored the relationship between a country's economic status and people's time use in 2016.
- Involved sourcing and combining data from various sources.
- Use **Excel**, **MongoDB**, **VLOOKUP**, etc. to analyze the impact of GDP.

Education App: Math/English Learning 2021

- Developed a web-based educational app using **ReactJS**.
- Organized project timeline with a **Gantt Char**t using Microsoft Project.
- Designed the UI in **Figma**, creating a cohesive learning platform.
- Implemented the Course Management for Tutors part, optimizing course administration.
- Developed **APIs** for Tutors part, improving data management.
- Ensured app quality through **testing** and code reviews.

Skills

Data analysis related skills:

- **SQL**: Able to write and understand SQL queries for data management.
- **XML**: Experienced in using XML for data formatting and transfer.
- **MongoDB** and **Robo3T**: Familiar with MongoDB and Robo3T for managing and querying data.
- **Tableau**: Skilled in using Tableau for data visualization and analysis.

Coding Language:

- Front-end: **ReactJS**, **JavaScript**
- Back-end: **Assembler Language**, **C#**, **Python**, **Java**
- Tools: **Postman**, **AWS**