Yingfei Shi

Phone: (+65) 82461599, Email: shi_yingfei@u.nus.edu
Address: Blk 435 Clementi Ave 3 #06-222 Singapore 120435

EDUCATION BACKGROUND

National University of Singapore (NUS) **Aug 2023 – Present**

- ***Major***: Information Systems and Analytics
- ***Degree Anticipated***: Doctor of Philosophy – ***GPA***: 4.67/5.0
- ***Research Interest***: Explainable AI Applications in Clinical Decision Support Systems

National University of Singapore (NUS) **Aug 2019 – May 2023**

- ***Major***: Business Analytics (Honors with Highest Distinction) & Economics (Second Major)
- ***Degree***: Bachelor of Science – ***GPA***: 4.81/5.0
- ***Award***: Dean's List (2020/2021 semester 1 & semester 2); Honor List of Student Tutors
- ***Certificate***: Analytics Techniques Knowledge Area (Distinction); Top Students for Application Systems Development for Business Analytics
- ***Relevant courses***: Programming Methodology, Business Analytics, Algorithms and Data Structure, Natural Language Processing, Machine Learning

ACADEMIC PROJECTS

Explainable Stock Price Prediction via Large Language Model **May 2024 – Present**

Supervisor: Vaibhav Rajan (Assistant Professor, NUS)

- Conducted an extensive literature review to outline the advancement and current state of CDSS in healthcare, particularly in cancer drug response prediction.
- Identified the critical gap between the current capabilities of Explainable AI (XAI) techniques and the clinical requirements for actionable, evidence-based explanations in cancer treatment decision-making.
- Employed a design science approach, leveraging publicly available knowledge bases and clinical evidence to support the interpretation of molecular alterations identified by the model.
- Proposed an innovative framework that integrates domain-specific knowledge and clinical evidence to generate clear, actionable, and clinically relevant explanations for AI model outputs, aimed at enhancing the interpretability and trustworthiness of AI-driven Clinical Decision Support Systems (CDSS) for cancer drug response prediction.

Explainable Stock Price Prediction via Large Language Model **Aug 2024 – Present**

Supervisor: Michael Shieh (Assistant Professor, NUS)

- Assembled and integrated financial news, historical trading volumes, and price trend data from reputable sources to create a comprehensive dataset for model training and testing.
- Developed a novel framework that incorporates state-of-the-art deep learning models including transformer architecture and large language models (LLMs) to enhance prediction accuracy and explanation depth.
- Advanced the model's ability to generate insightful and actionable stock movement predictions with a self-reflective mechanism that iteratively refines predictions and explanations, utilizing a unique combination of technical analysis and multi-level news assessment.

Cancer Drug Response Prediction Using Large Language Models **Jan 2024 – May 2024**

Supervisor: Wynne Hsu (Provost's Chair Professor, NUS)

- Leveraged comprehensive mutation profiles and drug characteristics from prominent databases and formulated the data into natural language descriptions to facilitate the processing by LLMs.
- Employed state-of-the-art LLMs adapted for biomedicine to predict cancer drug responses, using techniques such as fine-tuning and few-shot learning to enhance model accuracy and applicability in clinical settings.
- Conducted rigorous testing to compare baseline performance with enhanced versions through fine-tuning,

demonstrating that LLMs, especially when fine-tuned or using few-shot learning, show significant potential in improving the precision of cancer drug response predictions.

Cancer Drug Response Prediction **May 2022 – Aug 2023**

Supervisor: Vaibhav Rajan (Assistant Professor, NUS)

- Collected and pre-processed cancer data from various sources for model training and testing.
- Explored and experimented with deep learning models including ResNet, Transformer and other state-of-the-art models with clinically applicable cancer datasets to facilitate drug response prediction.
- Experimented with Natural Language Processing, Graph Representation Learning, and other techniques for cancer cell lines modelling to improve model performance for cancer drug response prediction under clinically applicable setting.

Indoor Scene Localization **Feb 2022 – May 2022**

Supervisor: Jian Lai Ng (Data Scientist, ST Engineering)

- Collected data for training from open sources via web scraping and fieldwork to serve as reference images in the database.
- Adapted existing scene localization models to assess the location of a single or multiple images of an indoor venue and outputted the readable signs/amenities/facilities within the query image and its geolocation candidates.
- Built a user-friendly front-end application by integrating the indoor and outdoor solutions with backend algorithms into an existing web application.

COVID-19 X-ray Image Prediction **Aug 2021 – Dec 2021**

Supervisor: Um Sungyong (Assistant Professor, NUS)

- Drew upon deep learning techniques to achieve high prediction accuracy on COVID-19 detection using X-ray images.
- Tested candidate models including manually structured CNN, CNN connected with XGBoost, and other pre-trained models on the data collected to assess their suitability for this problem.
- Developed the web-based application to address post-COVID hygiene requirements for quarantine hotels using HTML, CSS, and JavaScript.

Risk Analysis/Management **Aug 2021 – Dec 2021**

Supervisor: Kewei Huang (Associate Professor, NUS)

- Conducted financial analysis on selected technology mutual funds using qualitative methods to examine historical performance such as annual return, standard error of return, and Sharpe Ratio.
- Quantitatively analyzed the performance of mutual funds based on the selected utility function and performed portfolio optimization by adjusting the weights of holdings.
- Examined the effect of hedging techniques by comparing and studying the risk and return performance of selected portfolios.

INTERNSHIP EXPERIENCES

IIG Data Science Intern, GIC Private Limited, Singapore **May 2022 – Aug 2022**

Supervisor: Jingyuan Pan (Machine Learning Engineer)

- Performed data collection and data cleansing of Chinese national and regional policies and meeting notes data. Assisted in exploratory data analysis and data visualization to glean business insights.
- Explored and developed the sub-topic labelling and key phrase extraction algorithm based on word tokenizer, word dependency and part of speech tagging for the Chinese NLP policy analysis task pipeline.

Data & AI, Data Science Intern, SP Digital R&D, Singapore **Dec 2021 – May 2022**

Supervisor: Jingting Cher (Deputy Director, Data Science)

- Worked with the data science team on data pipeline development, energy forecasting, and anomaly detection model building to solve the energy management problem.
- Collaborated with data science professionals on the project of Water and Electricity Consumption Anomaly Detection, with responsibilities including:
  (1) Performed exploratory data analysis to analyze the electricity consumption data of clients to gain insights into the properties of time series data.

(2) Pre-processed the data and trained a machine learning model to detect anomalous consumption patterns in commercial and industrial buildings.
(3) Experimented with different state-of-the-art machine learning models and back tested the dataset to optimize the design of the model.
(4) Put the final model into production via the proper pipeline to create business value.

Data Analyst Intern, Beijing Kuaishou Technology, Shanghai, China **May 2021 – Jul 2021**
Supervisor: Jixuan Li (Director of Business Development Team)

- Facilitated the establishment of a problem and analyzed possible reasons/factors for anomalies by evaluating statistics of daily reports.
- Implemented data visualization by creating insightful daily dashboards of sales operation-related data with Tableau for the sales department, thus facilitating the tracking of the progress of sales plans.

ASSISTANTSHIP EXPERIENCE

Teaching Assistant, National University of Singapore **Dec 2023 – May 2024**
Teaching Module: Mining Web Data for Business Insights *Instructor: Wang Qiuhong (Senior Lecturer)*

- Preparing content for course projects and assignments.
- Grading students' assignments and project presentations and providing feedback and suggestions.
- Conducting group project consultations to help students with their problems.

Teaching Assistant, National University of Singapore **Aug 2020 – Aug 2023**
Teaching Module: Programming Methodology *Instructor: Leong Wai Kay (Senior Lecturer)*

- Teaching weekly tutorials to explain Python-related concepts and skills.
- Holding individual consultations or group Q&A sessions to help students with their problems.
- Grading students' submissions and providing crucial feedback.

COURSERA COURSES (self-study)

- Introduction to Artificial Intelligence with Python, Harvard University
- Computer Science: Programming with a Purpose, Princeton University
- Supervised Machine Learning: Regression and Classification, Stanford University
- Advanced Learning Algorithms, Stanford University
- Unsupervised Learning, Recommenders, Reinforcement Learning, Stanford University
- Tools for Data Science, IBM
- What is Data Science, IBM

TECHNICAL STRENGTH

- ***Skills & Tools***: Python, Java, R, SQL, STATA, MongoDB, HTML, CSS, JavaScript, Spark
- Proficient in performing exploratory data analysis, data pre-processing, feature selection, model selection, training, and evaluation.
- Familiar with Linear Regression, Logistic Regression, Clustering Algorithms including K-Means and K-NN, Decision Tree, Supporting Vector Machines (SVM), Naive Bayes, Random Forest, and Convolutional and Recurrent Neural Network.
- Solid command of machine-learning libraries such as Pandas, NumPy, Scikit-learn, Matplotlib and TensorFlow and PyTorch for deep learning.

LANGUAGES & INTERESTS

- ***Mandarin Chinese***: native speaker with excellent literary skills
- ***English***: fluent in reading, listening, speaking, and writing
- ***Interests***: piano, Chinese traditional painting, calligraphy, sketching, reading, hiking, and travelling