Soon Yung (Soung) Low

Personal website | GitHub | soung.low@outlook.com

WORK EXPERIENCE

Model Risk Data Scientist March 2023 - Present

NatWest Group, UK

- Performing validations of AI models across the bank, such as fraud prevention models and generative AI use cases
- Evaluating models in terms of modelling design, model performance, data quality, code implementation, and monitoring approaches
- Conducting tests and building challenger models to identify shortcomings in a model development lifecycle
- Assessing and communicating risks associated with the use of AI models to stakeholders and model developers
- Developing a Python-based LLM testing package that tests for accuracy, stability, toxicity, and vulnerabilities via red teaming
- Streamlining the model governance workflow using LLMs supported with retrieval augmented generation

Data Scientist Sep 2021 - Feb 2023

Amplifi Capital, UK

- Managed decision engines for loan and saving applications, such as designing and implementing decision rules for new products
- Implemented NPV model that works alongside a CatBoost model to find the optimal APR for loans
- Conducted analyses on credit bureau data, such as identifying temporally stable features for machine learning models
- Built prototypes of machine learning scorecard for credit scoring using tree-based methods e.g. LightGBM
- Documented changes in the decisioning process on Confluence and communicating dependencies to other teams
- Built a rejection funnel that shows the breakdown of declines by each decision rule to improve lending decision-making

Data Science Research Assistant July - Oct 2021

LSE Department of Government, UK

- Populated a dataset of legislative speeches of the Italian Chamber of Deputies from 1941 to 2018
- Scrapped PDFs for more than 5 million speeches from online archive and relevant legislature data from Wikipedia
- Converted and cleaned reports from PDF format into CSVs using OCR technique
- Sourced external metadata for speeches, including politicians, parties, and national elections

Machine Learning Intern for Nissan-FCU collaboration July - Aug 2020

FCU Artificial Intelligence Research Center, Taiwan

- Predicted the repurchasing rate of existing Nissan car owners using an XGBoost model
- Extracted features from maintenance data for time-series forecasting using sliding windows
- Conducted EDA on transaction data, maintenance records, and customer complaint data

EDUCATION

MSc Applied Social Data Science with Distinction Sep 2020 - Aug 2021

London School of Economics and Political Science (LSE), UK

- Grant awardee of LSE Saw Swee Hock Southeast Asia Centre Student Dissertation Fieldwork Grant
- Dissertation: “Ethnic Stereotypes in Malaysia: A quantitative measure based on word embeddings”
- Modules: Computer Programming, Applied Machine Learning, Quantitative Text Analysis, Distributed Computing for Big Data
- Data Science Society Project Developer for “Sentiment Analysis of S&P 500 Financial News”

B.A. Economics Sep 2017 - July 2020

Feng Chia University (FCU), Taiwan

- GPA: 4.0/4.0; Ranking: 1st/109
- Dissertation: “Sentiment Index of China’s Stock Market and its Causal Effect on Stock Indices”
- Professor and Chair Yao-Hsien Chien Memorial Scholarship Holder
- Modules: Statistics, Python Programming, Data Science using R, Econometrics, Empirical Methods

RESEARCH EXPERIENCE

Measuring Fairness in Financial Transaction Machine Learning Models (report) 2024

Conducted as a Data Study Group Researcher at the Alan Turing Institute, UK

- Collaborated with Mastercard's AI Governance to assess fairness of multi-class classification and regression models
- Operationalised fairness in terms of gender, ethnicity, age, and all intersectional subgroups of the aforementioned attributes
- Evaluated model biases using parity measures such as demographic parity, equal opportunity, and accuracy equality
- Explored bias mitigation approaches including adversarial de-biasing and exponentiated gradient post-processing method

Unveiling Racial Stereotypes in Malaysia using Word Embeddings (slides) 2023

Presented at the 5th International and Interdisciplinary Conference on Quantitative and Computational Analysis of Textual Data (COMPTEXT 2023)

- Provided a quantitative measure of racial stereotypes in Malaysia using word embeddings trained on new articles in 3 languages
- Proposed a method to calculate stereotype scores in terms of socioeconomic status, appearances, and personal qualities
- Identified words that are most associated with each race (i.e. Malay, Chinese, Indian) to understand mutual perceptions

'Wanita' in Parliaments: The attitude of Malaysian MPs towards women (slides) 2022

Presented at the 4th International and Interdisciplinary Conference on Quantitative and Computational Analysis of Textual Data (COMPTEXT 2022)

- Scraped 30-year Hansard of Malaysian parliamentary speeches (1991-2020) from the official portal
- Analysed temporal changes in the usage of gender labels in the Malaysian parliament in terms of frequencies and semantic meaning
- Quantified stereotypes against female MPs in personal traits and occupations using word embeddings

Investigating the Impact of the UK's Soft Drinks Industry Levy with Sainsbury's (report) 2022

Conducted as a Data Study Group Researcher at the Alan Turing Institute, UK

- Evaluated the impact of sugar tax on purchases of sugary drinks in collaboration with Sainsbury's
- Conducted exploratory data analysis and feature extraction using transaction data and customer demographics data
- Applied time series clustering technique to assess the levy effect on different groups of customers

The Hashtag Activism of Milk Tea Alliance on Twitter: A Mixed-Method Study 2021

Presented at the 12th Asian Conference on Media, Communication & Film 2021 (MediAsia)

- Studied the usage of social media (Twitter) among cloud activists for the Milk Tea Alliance in Asia
- Collected 89,091 tweets with the hashtag #milkteaalliance to assess the role of Twitter in the activism
- Conducted topic modelling and content analysis to identify key narratives and impact of the activism

SKILLS

Programming	Python (numpy, pandas, scikit-learn, transformers, autogen), R, SQL
Other Tools	AWS SageMaker, Git, Jira, Confluence, Microsoft Power BI
Languages	English (Advanced), Chinese (Native), Cantonese (Advanced), Malay (Advanced), French (Basic)