KOH SI XING

Email: ksxing8388@gmail.com   LinkedIn: https://sg.linkedin.com/in/kohsixing   Website: https://kohsixing.github.io

EDUCATION

National University of Singapore (NUS)   **Aug 2025 - Present**

Master of Science (M.Sc.) in Data Science and Machine Learning · *Part-time*

- Matriculated in Academic Year 2025/26

National University of Singapore (NUS)   **Aug 2020 - Dec 2023**

Bachelor of Science (B.Sc. Hons.) in Business Analytics · *Full-time*

- Honors: Honors (Distinction) - Second Class (Upper) Honors Equivalent
- Specialization in Financial Analytics

Singapore Polytechnic (SP)   **Apr 2014 - Mar 2017**

Diploma in Infocomm Security Management (with Merit) · *Full-time*

- GPA: 3.83 / 4.00
- Specialization in Forensics Investigation

EXPERIENCE

Research Assistant   **Jan 2024 - Jul 2025**

National University of Singapore (NUS) · *Full-time*

- **Image Classification Task:** Streamlined 1000+ participant recruitment and response collection for image classification using Prolific and Qualtrics APIs. Suggested label to participants using the GoogLeNet Inception V3 model, which participants may adjust. Engineered features based on participant responses, image difficulty, and image category; addressed data imbalance with SMOTE and class-weights. Trained a machine learning model to assess human-algorithm collaboration and integrated it into subsequent studies to generate real-time probability estimates. Developed and deployed R Shiny applications for image classification tasks, with selected applications incorporating the model to inform participants of their response accuracy.
- **Effects of Large Language Models (LLM):** Expanded upon my duties as a student research assistant, and studied the impact of LLMs on Reddit by comparing metrics like awards, number of comments, readability, and sentiment before and after ChatGPT's introduction. Applied traditional NLP methods (lemmatization, stemming) when using NLTK and Vader, alongside advanced HuggingFace models like DistilRoBERTa and DistilBERT.
- **Predicting Inventory Wastage:** Engineered features like time lags, product categories, and historical adjustments to build a machine learning model, predicting the adjustment quality made by store managers, which reduced wastage for perishables due to unnecessary adjustments. Optimized the model's performance using parallel programming on a High Performance Computing (HPC) cluster when conducting grid search and hyper-parameter tuning.

Student Research Assistant   **Sep 2023 - Nov 2023**

National University of Singapore (NUS) · *Part-time*

- **Effects of LLMs:** Conducted literature reviews to explore the impact of LLMs on online communities like Reddit and Stack Overflow. Analyzed subreddits and communities with frequent discussions to investigate changes in language and human behavior before and after ChatGPT's inception. Automated data collection workflows using GitHub Actions to schedule scraping from selected subreddits, ensuring consistent data collection for the project.
- **Human-Algorithm Collaboration:** Conducted literature reviews to examine existing experiments on human-algorithm collaboration, identifying their research methodologies and objectives to enhance empirical studies in this domain.

Business Analytics Intern   **Jan 2023 - Jul 2023**

The Development Bank of Singapore (DBS) Limited · *Internship*

- **Data Collection:** Automated the collection of employee reviews from DBS and selected companies to analyze employee experience within DBS relative to industry standards. Distributed Qualtrics surveys to capture employee sentiments and leveraged Qualtrics' built-in dashboards for quick analysis of survey results.
- **Data Analysis:** Utilized Python for data cleaning and exploratory data analysis (EDA) on scraped data. Applied moving average charts for trend analysis and linear regression for causal inference, visualizing charts with Tableau and Python Plotly.

- **NLP Techniques:** Processed and engineered features from a dataset of over 2 million employee verbatims. Fine-tuned the DistilRoBERTa model to classify reviews into relevant sentiment categories, achieving an F1 score of 0.85, which demonstrated high accuracy in verbatim classification.
- **Stakeholder Presentation:** Presented actionable insights derived from data analysis and NLP to the Managing Director and senior managers of the Transformation Group. Visualizations such as charts from Tableau and Plotly were used to support recommendations such as enhancing work culture and providing office welfare (e.g. vending machines or increased shuttle bus frequencies in less accessible locations), were incorporated to improve overall employee satisfaction.

Associate Research Engineer   Aug 2019 - Aug 2020
Trustwave Holdings · *Full-time*
- **Data Collection:** Gathered data from online malware reports and scraped cybersecurity sources to build a corpus for NLP-based malware indicator classification.
- **Data Analysis:** Enhanced report processing script to extract malware indicators from PDF/HTML files. Conducted correlation analysis to remove redundant features and weighted key features to improve model performance.
- **Model Building:** Trained a multi-label classification model using extracted features, visualized via spider charts to show the likelihood of each malware category. The model achieved an ROC score of at least 0.88. Applied 10-fold cross-validation to prevent overfitting and ensure model robustness.
- **Prototype Development:** Developed an automated system using Plotly Dash and Python, to ingest and process malware reports from various cybersecurity tools. The integrated machine learning model predicts relevant malware categories, displaying results in a user-friendly interface with summary reports and severity levels.

Cyber Security Intern / Associate Engineer   Sep 2016 - Jun 2017
Singapore Telecommunications (SingTel) Limited · *Internship*
- **Proof of Concept (POC):** Evaluated various cybersecurity technologies, such as Cloud Access Security Brokers (CASB) and Deception Technologies, to support strategic partnerships and recommend solutions to clients.
- **Anomaly Detection:** Conducted EDA on network traffic data using R to identify indicators of potential cyberattacks; Presented findings to supervisors and the Director of Technology & Operations.

Skills
- Programming Skills: Java, Python, R
- General IT Skills: Computer Networking, Database (SQL, MongoDB), Web Development, Git
- Data Science & ML: Scikit-learn, Imblearn, Tensorflow, Dplyr, HPC Cluster, HuggingFace
- Cybersecurity: Kali Linux, REMnux SIFT

Awards
- Top Students for BT2102 Data Management and Visualization (Associated with NUS)
- SingTel Cybersecurity Cadet Scholarship (Associated with SP)
- Singapore Polytechnic Scholarship (Associated with SP)
- Director Honor Roll for AY2015/16 (Associated with SP)

Certifications

Financial Analytics Specialization
Certificate of Merit
National University of Singapore

Higher Mathematics
Certificate of Completion
Singapore Polytechnic

Organizational Systems Wireless Auditor
Wireless Auditing & Pentesting Certification
ThinkSECURE

Organizational Systems Security Analyst
Enterprise Security Certification
ThinkSECURE

Mobile Device Forensics
Certificate of Completion
Paraben Corporation