Nicholas WAN

DATA SCIENTIST

nicholaswan@live.com.sg   http://www.nicholaswan.me/   https://github.com/nicholas-wan

Education

National University of Singapore   *Singapore*

BS.C BUSINESS ANALYTICS (HONS), HIGHEST DISTINCTION   *Aug. 2017 – Dec 2020 · 3.5 yrs*

• NUS Merit Scholarship
• Cumulative Average Point: 4.66/ 5.00
• NUS Overseas Colleges (Silicon Valley)

Stanford University

NON GRADUATING OPTION   *Aug. 2019 – Apr 2020 · 9 mos*

• Mining Massive Datasets – Grade Obtained: A
  Topics Include: Frequent itemsets and Association rules, Near Neighbor Search in High Dimensional Data, Locality Sensitive Hashing (LSH), Dimensionality reduction, Recommendation Systems, Clustering, Link Analysis, Large-scale Supervised Machine Learning, Data streams, Mining the Web for Structured Data, Web Advertising.

Skills

Programming	Python · R · HTML/CSS · Java · Javascript
Database	MySQL · MongoDB
Frameworks	Spark · VueJS
CI	Git

Experience

Defence Science and Technology Agency   *Singapore*

ENGINEER (CYBER AI)   *Jan 2021 – Present· 1 yr 5 mos*

• AI and Machine Learning solutions to defend against Cybersecurity threats.
• Developed models to detect Phishing based on URLs, logos, and Website Screenshots, which successfully detected Phishing websites and achieved an F1 score of 0.83 on a public validation dataset from Phishtank. Automated the alerts in the forms of PDF reports to be sent in emails to stakeholders to take action.
• Created a Tech Showcase (Interactive Website) on the topic of <Adversarial AI Attacks on Smart Home devices> for BrainHack 2021, an annual DSTA organised event. The showcase was voted by participants as the top showcase out of 27 showcases that were created by other departments in DSTA.

Shopee   *Singapore*

DATA SCIENCE INTERN   *Apr 2020 – Nov 2020 · 8 mos*

• Trained NLP models to perform multiclass classification on product titles and descriptions to predict product attributes.
• Implemented NLP in multiple languages including English, Chinese, Thai, Indonesian and Vietnamese.

The League   *San Francisco*

DATA SCIENCE INTERN   *Aug 2019 – Apr 2020· 9 mos*

• Created and productionized a Machine Learning model to predict catfishers and bad users leading to the customer support team saving 30% of time reviewing profiles and improving the trust and safety in the community.
• Set up an alert system for over 20 key company metrics such as Revenue, DAUs, Churn, using Periscope / Sisense to notify stakeholders through Slack.
• Performed Statistical modelling through regression models such as Diff in Diff, IV Regression, Fixed Effect models to advise product decisions.
• Conducted AB Testing experiments with power analysis and statistical significance tests to inform and influence product decisions, which included several product features.
• Experience with creating, scheduling, monitoring and debugging jobs using Airflow.

National University Hospital   *Singapore*

DATA SCIENCE INTERN   *Dec. 2018 – Jun 2019· 8 mos*

• Created a model for predicting patient readmission within a 30 day period, using Natural Language Processing and Neural Networks, and Random Forests.
• Worked on generating a realistic synthetic dataset to be released to the public and used in Hackathons.
• Automated the process of extracting information from Free-Text form for Gastro-Translational Research.

Certifications

Advanced Natural Language Processing and Temporal Sequence Processing   Red Dragon AI

• Topics include: Named Entity Recognition, RNNs, Embeddings, Seq2Seq, Neural Machine Translation, Text Generation, BERT, Transformers, Intent Classification, Chatbots.