

Chan Khai Ern, Edwin

Contact Number: +6592979020 | Email: edwin.ke.chan@gmail.com

Education

National University of Singapore

Aug 2012 – July 2017

- Bachelor of Science in Life Sciences (Honours with Highest Distinction)
- Double Minors in Computer Science and Physics
- Thesis Project: Deep Learning and Bayesian Inference in Variant Calling
- Selected into University Scholars Programme (USP), a multidisciplinary academic programme that accepts only 3% of each undergraduate cohort

Work Experience

Ministry of Defence, Singapore

Data Scientist

Oct 2017 - Present

Research and Development

- Applied statistical and machine learning techniques to solve complex problems in national defence and cybersecurity, such as network threat/anomaly detection and fake news analytics.
- Explored and analysed large multimodal datasets at terabyte scale using unsupervised learning techniques like gaussian mixture models to develop new use-cases and discover potential areas of research
- Deployed a dockerised end-to-end analytics application that automated 14 manual analyst processes, enabling a single data analyst to monitor an entire team's worth of results and queries. The application is currently being used in real operations.

Software Engineering

- Deployed candidate end to end machine learning models, including building ETL data streams, Django backend systems and visualization suites to support fast feedback loops from users
- Develop and maintain internal python software libraries and R&D databases to increase overall team efficiency and optimise internal R&D tech stack, including reviewing PRs and ensuring coding standards
- Set up devops tools like gitlab, atlassian suite (JIRA, Confluence, Mattermost) and containerized CI/CD pipelines for data science team in order to enable modern development workflows within air-gapped networks

Institute of Water Policy, National University of Singapore

Research Assistant

May 2015 - May 2016

Statistical Analysis for Public Policy

- Performed statistical analysis on hotel water consumption datasets to recommend changes to water pricing policy
- Submitted a technical report to the Singapore Public Utilities Board analyzing the determinants of hotel water consumption

Achievements

Valedictorian, NUS Life Sciences 2017

- Awarded for outstanding academic achievement and contributions to student community

University Scholars Programme Presidents Honours Roll 2016

- Awarded for academic excellence and strong leadership in co-curricular activities

NUS Undergraduate (Merit) Scholarship 2013

- Fully tenured scholarship for undergraduate studies
- Awarded for exceptional academic achievements and strong leadership potential

Projects

Deep Learning and Bayesian Classification for Variant Calling

- Designed and built a deep learning network to improve variant calling precision, and designed Bayesian networks for probabilistic ranking of functional genes.
- Research Citation : **Chan E.**, Lakshmanan, M., Tan, T. W., Ban, K. (2017, August). Integrated Deep Learning and Bayesian Analysis for Prioritization of Candidate Cancer Genes from Next Generation Sequencing Data. Poster presented at the 4th National Cancer Institute of Singapore (NCIS) Annual Research Meeting, Singapore.

pandas-profiling

- Active contributor to pandas-profiling, a python data profiling library with over one million downloads and tens of thousands of users.
- Leading an international group of data scientists to design and build profiling tools for spark dataframes, with over six thousand LOC contribution.

pydistinct

- Creator of pydistinct, a python library that implements 14 statistical estimators to predict the cardinality of a population from a sample.
- The library has been downloaded over 4000 times and has been cited in the paper “On estimating the number of flows.” by Bruce Spang in the Stanford Workshop on Buffer Sizing.

Leadership

42nd Life Sciences Society Executive Committee

President

Aug 2013 - Aug 2014

- Initiated a wide range of programmes for professional development of students, including the creation of a peer mentorship programme with over 150 freshmen participating.
- Spearheaded various activities to contribute back to the community, such as a bimonthly volunteering program with Melrose Home, a home for children from vulnerable backgrounds.

Skills

Languages

Python, Bash, SQL, Java

Proficient technologies and libraries

Apache Spark, scikit-learn, pytorch, keras, pandas, flask, django, docker, streamlit

Certifications

- Algorithms by Stanford University on Coursera. Apr 21, 2020
- Statistical Pattern Recognition by National University of Singapore. Dec, 2018.
- Neural Networks and Deep Learning by Stanford University on Coursera. Oct 14, 2017