

Chengyi (Jeff) Chen

SG: +65 98586602 | jeffchenchengyi@gmail.com | <https://jeffchenchengyi.github.io>

SKILLS

Technical Skills: Python (Sklearn, Scipy, Numpy, Cvxpy, Matplotlib, PyTorch, Pyro, PyMC3, Tensorflow, PySpark) | SQL

WORK & LEADERSHIP EXPERIENCE

Gojek Singapore

Singapore, Singapore

Data Science Intern, Pricing Team

May 2020 – Aug 2020

- Dynamic / Surge Pricing | Technologies used: numpy, tensorflow, cvxopt, cvxpy
 - Contextual Bandits: Off-Policy Evaluation and Error Bound Calculation
 - Research on off-policy value estimators:
 - Bias, Variance, Mean-Squared Error Analysis of 1. Inverse Propensity Scoring (IPS), 2. Doubly Robust, 3. Self-Normalized IPS, and 4. Maximum Empirical Likelihood estimation.
 - Implemented and compared error bounds for the IPS estimator such as t-distribution, asymptotic gaussian, clopper-pearson, bootstrapping, and ones derived from Hoeffding and Bernstein inequalities
 - Investigated convergence of off-policy value estimates of the target policy to the actual value

Shopee Singapore

Singapore, Singapore

Data Science Intern, Marketing Science

December 2019 – May 2020

- Churn Prediction | Technologies used: pyspark, pyspark sql, pytorch, pyro, shap, sklearn, plotly
 - Model Performance Tracking and Explanation:
 - Presented contribution of features used in LightGBM models to marketing managers and key stakeholders using SHAP
 - Used Plotly to generate animations displaying incumbent model's performance across all 7 regional markets
 - Model Exploration and Feature Engineering:
 - Explored other pyspark ml, H2O's AutoML binary classifiers and MMLSpark survival models
 - Reformulated Churn Prediction into a time series regression problem instead of binary classification and developed a PyTorch Sequence2Churn model to predict time to churn
 - Developed end-to-end feature engineering pipeline to process raw data from parquet files on Hadoop, producing both static and time series features
- Voucher Sensitivity | Technologies used: causallift, pyro
 - Researched on amount of uplift generated using different vouchers and implemented code to estimate the Conditional Average Treatment Effect using Inverse Propensity Weighting / Scoring

EDUCATION

University of Southern California (USC)

Los Angeles, California

M.Sc. in Analytics and B.Sc. in Computer Science Business Administration

December 2021

Grad GPA: 4.00 / 4.00 | UGrad GPA: 3.84 / 4.00 | SAT: 1550

PROJECTS

Evolving FPGA Research with Center for AI in Society's Student Branch (CAIS++)

September 2019 – Present

- Evolving Field Programmable Gate Array (FPGA) circuit configurations to become universal function approximators competitive with neural networks using genetic algorithms and evolutionary strategies such as novelty-search with a variety of distance metrics (e.g. wasserstein) and multi-objective optimization.

Exploring Housing Prices in Singapore

May 2019 – August 2019

- Scraped www.99.co (Singapore Property Portal) for property features and transaction history using BeautifulSoup.
- Performed Clustering (K-means) and Regression (Random Forest) analysis on the data, followed by a brief exploration of the most popular condominiums in Singapore.

Education Deserts Research [CSCI499: Artificial Intelligence for Social Good]

January 2019 – May 2019

- Cleaned US Census data (Tract level), Geospatial data (Shape files for US States), and university coordinates to find education deserts (tracts with no universities within a 25-mile radius) and used random subsampling to balance dataset used for building education desert classifier and performed recursive feature elimination to discern important features of an education desert (Population Density / square mile, Median Gross Rent, ...).

Udacity Data Scientist Online Nanodegree Program

January 2019 – August 2019

- Completed projects ranging from building Recommendation Systems using Matrix Factorization techniques (Singular Value Decomposition) for Collaborative Filtering to predicting Customer Churn with the PySpark API.