Chengyi (Jeff) Chen

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

SKILLS

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

WORK & LEADERSHIP EXPERIENCE

Plutus Mazu — **Singapore, Singapore**
Data Scientist — June 2021 – Dec 2021

- Quant Team | Technologies used: sklearn, imblearn, optuna, talib, plotly, dash, scipy, numpy
  - Machine Learning Wrapper over Trading Strategies
    - Extended sklearn library for proprietary trading strategies
    - Implemented novel machine learning algorithms such as a Quantile Ensembler and Residual Regressor model with appropriate cross-validation techniques.
    - Built an end-to-end inference / model selection pipeline from scratch, including data preprocessing, feature engineering, novel dimensionality reduction methods, novel feature selection methods, hyperparameter optimization, user interface for choosing "best" model from the pareto-frontier of multi-objective optimization problems and converting it into deployable, regularly retrained models

Gojek Singapore — **Singapore, Singapore**
Data Science Intern, Pricing Team — May 2020 – Aug 2020

- Dynamic / Surge Pricing Team | 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 — Dec 2019 – May 2020

- Churn Prediction Team | 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

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.