

Marine (Meng) Lin

San Francisco, CA • (424) 535-6298 • marinelin33@gmail.com • github.com/marinelin

Education

University of San Francisco

June 2020

M.S. in Data Science

Coursework: Machine Learning, Deep Learning, Distributed Computing (Spark), Data Structure and Algorithms, Time Series Analysis, Relational Databases, A/B testing, Product Analytics, Data Ethics

University of California, Los Angeles

Sept 2015 - June 2019

B.S. in Statistics, Minor in Japanese

Technical Skills

Language and software: Python, R, SQL, Spark, Pytorch, Conda, H2O, Flask, AWS EC2 & S3, SAS, Stata, Tableau, Git

Work Experience

Reputation.com

Redwood City, CA

Data Science Intern

Nov 2019 - Present

- Matched over 70k U.S. business locations with 99% F1 score using fuzzy matching algorithm and deep learning models inside a docker container
- Classified COVID-19 related customer review topics using LDA and n-gram models in Python
- Analyzed the trend of COVID-19 reviews and presented the results to the business stakeholders

605 TV

Los Angeles, CA

Data Analytics Intern

July 2018 - Sept 2018

- Wrote SQL queries in Redshift and supported engineering team with data insights for new business models
- Validated advertisement viewership data for *Spectrum* and reduced the mismatching by 5%
- Provided marketing insights of viewership to executive and other internal teams through building logics for viewership segment definition using R's dplyr package
- Translated large-scale (>150k) household information across the U.S. into business deliverables using R

Uber China

Beijing, China

Data Analytics Intern

July 2017 - Sept 2017

- Performed ad-hoc data analysis through SQL which resulted in a 3.5% reduction of miscalculation of the passenger pickup rate
- Conducted A/B testing to optimize promotion effectiveness and increased user engagement by 2%
- Mined complex customer behavioral data with Tableau and visualized the results which changed the business plan and led to an increase of Uber's revenue by 30% in Chongqing City

Data Science Projects

Basketball Pickup Community Product Analysis

Spring 2020

- Led product development strategy by setting up the benchmarks for the MVP
- Built a web app with Flask and Bootstrap deployed on AWS Elastic Beanstalk
- Implemented UI design for the web app using HTML, JS, and CSS

Patient Condition Prediction Kaggle Project

Spring 2020

- Predicted short-term outcomes in critically ill patients with 90% accuracy, using XGBoost Regressor and Linear NN on Pytorch and ranked top 10% in the competition

Bay Area BikeShare Analysis

Winter 2019

- Increased accuracy by 15% by cleaning and analyzing 4GB of data with Spark SQL and Spark Machine Learning
- Built Random Forest, gradient boosting (xgboost), and H2O models with different AWS EMR clusters and predicted the bike return rate based on weather, location, and population with RMSE of 2.38
- Identified and visualized the trend in the variables by their geolocation with matplotlib and mapbox API

Twitter Sentiment Analysis

Fall 2019

- Performed sentiment analysis using the vaderSentiment library with data pulled from the tweepy wrapper and launched a web server at AWS that can display tweets with different colors based on the sentiment analysis