Beiming Liu

Data Science | Machine Learning | Predictive Modeling | Statistical Analysis | Data Mining | A/B tests

415-770-5088 | San Francisco, CA 94105 | ben.liu2013@gmail.com | beimingliu.me

Programming: Python (Scikit-Learn, Pandas, NumPy, XGBoost, PyTorch), Matplotlib, R
Database: SQL (Amazon Redshift), NoSQL (MongoDB)
Distributed Computing: Spark(spark.ml), AWS (EC2, S3)

WORK EXPERIENCE

Tally Technologies, Inc. | Data Science Intern | San Francisco, CA Oct 2017 – Present

- Classified applicants loan risk using tree-based machine learning methods and feature extraction, outperformed the previous risk model by 17%. The model is projected to save Tally $200K annually from delinquencies once in production
- Developed a statistical model and NLP technique to impute missing credit transactions categories and to correct misclassed categories, achieved 96.4% accuracy on the validation set
- Produced a data pipeline from users' credit transactions to predict their annual income using regression and time series models. Identified users whose claimed income was inconsistent with the model, expected to reduce manual income verification by 20%

Hfax.com | Data Science Intern | Beijing, China May 2017 - Jul 2017

- Analyzed $1.6B (10B RMB) of investments across half a million transactions: visualized user engagement and predicted business growth. Presented key findings and recommendations to the management team
- Discovered groups of similar users on the platform using unsupervised/semi-supervised machine learning algorithms to identify investment attributes, provided suggestions on marketing strategies
- Produced a detailed user profile for each investor based on attributes: engagement, loyalty, influence, (short-term) profitability, and LTV

EDUCATION

University of San Francisco | M.S. Data Science | San Francisco Jul 2017 - Expected Jun 2018

Coursework in A/B tests (experiments design), Machine Learning, Distributed Computing, Natural Language Processing (NLP), Linear Regression, SQL, Time Series, Statistical Modelling and App Development

Imperial College London | MSc Financial Math, BSc Mathematics | London Oct 2011 - Jun 2016

PROJECTS

Predictive Analysis: What's the story behind 17 million Citi Bike Trips in 2017?

- Predicted duration for each trip in NYC, using weather data and distance estimation (from GPS data)
- Implemented regression and time series models to predict total daily trips and achieved R^2 of 0.82

NLP Sentiment: What does Trump say and how the stock market reacts?

- Scraped Trump's Tweet since 2017 and performed NLP technique to calculate the sentiment scores
- Visualized the correlation between the sentiment and the stock market performance

Web-application: Web-application: LEGOIT.US

- Develop a web application that transforms photos into a Lego set with Instagram-like filters
- Designed the APP from scratch and managed a team of 8 on data pipeline, algorithm and front-end

Linear Regression Analysis: House Price Prediction

- Constructed a regression model to predict house price, validated model fit and normality assumption, using OLS, Ridge, Lasso and Elastic Net techniques

Deep-learning: Autonomous Driving on a simulated track

- Implemented model architecture from NVIDIA Self-Driving Car paper
- Fine-tuned parameters, reduced overfitting, and created autonomy on the simulated track