Beiming Liu

Actively looking for Data Scientist positions in the Bay Area

Predictive Modeling | NLP | Machine Learning | Time Series | Statistical Analysis | A/B tests

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

Data Science Intern — Oct 2017 – Present
Tally Technologies, Inc. | Series B, Venture backed Fintech startup — San Francisco, CA
- 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
- Designed and architected an end-to-end machine learning pipeline to fix missing and misclassified bank credit transaction categories (payroll, credit etc.) using a proprietary statistical model and NLP, achieved 95.4% accuracy
- Implemented statistical models to verify user's income using transactional data. Worked with the product, business and customers teams cross-functionally to verify users income using regression and to double the credit limits for low-risk customer groups

Data Science Intern — May 2017 - Jul 2017
Hfax.com | An Fintech Unicorn completed securitization volume over $22B — Beijing, China
- 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

SIDE PROJECTS AND INTEREST

Predictive Analysis: What's the story behind 17 million Citi Bike Trips in 2017? [link]
- Predicted duration of each trip in 2017 in NYC using gradient boosting method with 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? [link]
- Visualized the correlation between Trump's tweets sentiment and the stock market performance by scraping his tweets since 2017 and performed NLP technique (Vader sentiment) to calculate the sentiment scores

Web-Development: LEGOIT.US
- Managed a team of 8 to develop a web application that transforms photos into a LEGO set with filters
- Generated user manuals and optimized cost by web scraping real-time data from third-party sellers

Linear Regression Analysis: House Price Prediction [link]
- Predict house price using a regression model with OLS, Ridge, Lasso and reduced a business report

Deep-learning: Autonomous Driving on a simulated track
- Implemented model architecture from NVIDIA Self-Driving Car paper using self-collected images
- Fine-tuned parameters, reduced overfitting, and created autonomy on the simulated track

Interest: Traveled to 26 countries over 150 cities worldwide and 30 out of 34 provinces in China