Beiming Liu

Predictive Modeling | NLP | Machine Learning | Dimension Reduction | Data Analysis | Clustering

415-770-5088 | San Francisco, CA 94105 | ben.liu2013@gmail.com | www.beimingliu.me | github.com/benliu91

DATA SCIENCE EXPERIENCE

Data Scientist Intern Oct 2017 – Jun 2018
Tally Technologies, Inc. | Series B, Venture backed Fintech startup San Francisco, CA

Credit Model: Machine Learning

- Classified applicants loan risk using tree-based machine learning methods and feature extraction, outperformed the previous risk model (the expected loss rate) by 7%. The model is projected to save Tally $100K annually from delinquencies once in production

Transaction Categorization: NLP, ETL

- Designed and architected an end-to-end machine learning pipeline to fix missing and misclassified credit transaction categories (payroll, deposit etc.) using a statistical model and NLP, achieved 95.4% accuracy

User Income Prediction, Loan Modification: Time-Series

- 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 Scientist Intern May 2017 - Jul 2017
Hfax.com | A Fintech Unicorn completed securitization volume over $22B Beijing, China

Statistical Inference, Customer Segmentation

- 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

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

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

- Predicted the individual ETA and aggregated daily demand 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

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

SKILLS

Programming: Python (Scikit-Learn, Pandas, NumPy, XGBoost, Matplotlib), Tableau, Bash, Command-Line, Git, R
Database: SQL (Amazon Redshift, PostgreSQL), NoSQL (MongoDB)
Distributed Computing: Spark(spark.ml), AWS (EC2, S3), Unix Environment, TensorFlow, Keras