# Fuyu Zou

fuyu_zou@brown.edu | (401) 688-8145 | https://freyazou.github.io

## Work Experience

**Machine Learning Engineer Intern** June ~ August 2018
BrainCo Inc. Somerville, MA

- Down-sampled the large-scare EEG data with Map-Reduce programming model with Spark
- Developed real-time classifier for sleep stages based on recent econometrics research papers
- Implemented Neural Nets onto selected features and achieved state-of-art 93.1% high test accuracy
- Designed experiments to optimize model on new subjects, compiled the results into a presentation

**Research Assistant** January 2018 ~ March 2018
Center of Gerontology and Healthcare Research Providence, RI

- Developed hierarchical generalized linear model to predict re-admission rates for patients with CVD
- Checked the validity of the dataset and conducted exploratory data analysis based on research paper
- Utilized cross validation to optimize the model, and simulated interval estimates using bootstrap
- Summarized the final model and compiled research results into a presentation to other researchers

**Data Analyst Intern** July 2016 ~ December 2016
PayPal Holdings, Inc. Shanghai, China

- Collected, collated and carried out complex data analysis of in support of client managers' requests
- Improved analysis efficiency by 60% by establishing models that classify clients on their characteristic
- Minimized the time spent on monthly reports by synchronizing data in Excel and PowerPoint via VB
- Reduced the client churn rate by 10% through analyzing the best advertise timing to retain clients

## Selected Projects

**Prediction of Usefulness of the Yelp Reviews (NLP)**

- Parsed restaurants information from Yelp Dataset and implemented relational database using SQL
- Preprocessed text and utilized GloVe and n-gram methods to convert text into word representations
- Built ensemble supervised learning models (bidirectional-LSTM, CNN) to predict the useful votes
- Achieved highest test AUC of 94.5% in classification and visualized key words in a useful review

**Kaggle competition: New York City Taxi Trip Duration**

- Implemented Elastic net, generalized additive model (GAN), and LGBM to predict taxi trip duration
- Visualized the data in Plotly to do exploratory data analysis and potential spurious trips identification
- Utilized K-Means clustering algorithm to analyze the frequency of pick-up and drop-off spots on map
- Developed a LGBM based model that achieved 0.425 Root mean square logarithmic error on test dataset

## Education Background

**Brown University - M.S. in Biostatistics**
Providence, RI | GPA: 3.9/4.0 2017 ~ May 2019

- **Relevant courses**: Machine Learning, Deep Learning, Introduction to Algorithms, Numerical Optimization
- **Teaching Assistant**: Intro to Machine learning, Probability and Stats, Applied Regression Analysis

**Shanghai University of Finance and Economics**
Shanghai, China | GPA: 3.5/4.0 2013 ~ 2017

## Skills

**Programming**: Python, SQL, R, SAS, Julia
**Visualization**: Tableau, Plotly
**Machine Learning**: Tensorflow, Keras, Scikit-learn
**Data Infrastructure**: Spark, Google Cloud Platform