

JADE YUN

(832) 248-9433 | San Francisco, CA 94105 | yyun1988@gmail.com | [linkedin.com/in/jadeyun](https://www.linkedin.com/in/jadeyun) | github.com/yuyun2

EDUCATION

M.S. in Data Science, University of San Francisco 07/2017 – 06/2018

- Coursework: Machine Learning, Computational Statistics, Relational/NoSQL Databases, Data Visualization, Experimental Design (A/B Testing), Statistical Modeling, and Text mining.

B.S. in Biology, Minor in Kinesiology, University of Houston 08/2012 – 05/2017

- Graduated with Distinction.
- Coursework: Statistics for Science, Calculus, and Biostatistics.

WORK EXPERIENCE

SF17 (former Simpatica Medicine) Data Science Intern San Francisco, CA
10/2017 – 06/2018

- Speed-up of autoimmune diseases diagnosis by building a decision tree based model to predict patient's disease state using genetic information in Python (Random forest, XGBoost).
- Discovered patterns of important feature interactions that underlie biological pathways to help personalize medication.
- Increased the AUC score by 10% by performing feature engineering and tuning the model to optimize the algorithm's performance.
- Identified and quantified the trade-off between numbers of genes analyzed and model accuracy to simplify model complexity.
- Produced visualizations to characterize the relationship of features to aid discovering root causes of disease using Tableau.

DATA SCIENCE PROJECT

Risk Detection for E-Retailers (JD Finance Global Data Challenge)

- Predicted transaction risk for a Chinese e-retailer, JD.com, using machine learning techniques in Python (Logistic regression, Random forest, LightGBM).
- Identified 70% of fraudulent transactions with a 92% accuracy by combining login and purchase data extracted from a relational database (PostgreSQL).
- Ranked No. 1 in North America Division, presented the solution to executives from JD.com at JDD summit.

Health Insurance Claims Prediction

- Predicted the probability that a patient files an insurance claim for type II diabetes with an AUC score of 0.795 in Python (Logistic regression, Random Forest, XGBoost).
- Extracted relevant information from a collection of semi-structured JSON blobs containing medical records for 100K+ patients, and created a set of structured relational tables (SQLite).
- Generated 200+ features using SQL and identified the top predictive feature based on the trained model.

Legoit

- Created a web application that allows users to transform any images into a Lego set and generated PDF instructions to build the Lego image as well as providing links to purchase the necessary bricks.
- Designed and developed front-end for legoit.us using HTML, CSS, and Flask, and set up google analytics to monitor user behavior and web traffic.

Mobile App Advertisement Click-to-Download Prediction

- Detected fraudulent mobile clicks with an AUC score of 0.978 using decision tree based learning algorithms (Random Forest, LightGBM).
- Optimized memory efficiency when analyzing data consists of 200 million clicks by applying MapReduce, batch processing, and hashing techniques.
- Performed extensive feature engineering using click activities and device information to uncover fraud signals using Python.

Bay Area Bike Share Demand

- Built a distributed data storage system on AWS using MongoDB and S3.
- Predicted daily bike demand with a decision tree based model (Random forest) using PySpark and Spark ML library.

SKILLS

Python (Scikit-Learn, Pandas, Numpy, SciPy, Matplotlib, Plotly), R, SQL, PyTorch, Shell, Spark, Tableau, AWS, and Git