Min Gu (Min) Jo

(510) 365-4988 | mgj9993@gmail.com | Berkeley, CA
Personal Website : mingujo.github.io
GitHub : github.com/mingujo
LinkedIn : linkedin.com/in/mingujo

SKILLS

Languages*: Python, SQL, Java, R, Ruby*
Frameworks & Libraries*: Apache Spark, AWS(EMR, S3), Django, TensorFlow, Redis, Celery, ElasticSearch*
Databases & Platforms*: PostgreSQL, Amazon Redshift, Heroku, Airflow*
Technical Concentrations*: SaaS • REST • ETL (Big data processing) • ML Analytics • NLP • Database*

EDUCATION

University of California, Berkeley Class of 2016
B.A. in Computer Science, Statistics, and Economics
Cumulative GPA: 3.42/4.00

WORK EXPERIENCE

June 2016 - Present

Software Engineer - *Leadgenius, Berkeley CA*

- Built an automatic sales outreach email reply labeling application to serve 50+ customers from scratch
  - Classified email replies in 7 different labels by applying NLP algorithms
  - Alerted customers of positive replies in real-time by deploying a trained model on a server
    → Overall Accuracy: 92.8% (Accuracy on positive reply: 89.2%)
    → Increased returning visitor rate of the company's outreach product by 72%

  Used: TensorFlow, Redis, Celery
- Developed an ETL processing pipeline to store 25M+ U.S. based company and 30M+ professional data from a variety of sources
- Designed a data reduplication algorithm to store and index merged data
- Implemented a distributed search for database of 25M+ company data using ElasticSearch

  Used: Python Spark, AWS, Airflow, Django, ElasticSearch

Jan 2016 - Dec 2016

Research Assistant - *Berkeley Institute of Data Science, Berkeley CA*

- Implemented data ingestion pipeline of 5000+ movie review data into S3 via web scraping framework
- Trained binary sentimental classification model to label user reviews using Bag of Words model

PROJECT EXPERIENCE

Fall 2016

UC Berkeley Family Housing - Open House Scheduling Calendar

- Developed a calendar web app using Rails to serve 15+ UCB housing staffs for coordinating open house schedules with 30+ resident assistants. Automated email notification for any schedule changes.

Spring 2016

Kaggle Challenge: Rossmann Drugstore Store Sales Prediction

- Used and compared 3 different machine learning algorithms to forecast drugstore daily sales: multivariate linear regression, random forest regression, and gradient boosting with regression trees

Fall 2015

Kaggle Challenge: Bag of Words Meets Bags of Popcorn

- Applied text analysis (NLP) methods of TFIDF vectorizer and Google's word2vec on iMDb movie reviews to perform sentiment analysis (96% accuracy | top 11th percentile when submitted)

Spring 2013

Detection of Activated Brain Regions Under Mixed Gamble Task

- Investigated the relationship between brain activity and behavior of the subjects towards the 50/50 gambling situations using a whole-brain robust regression analysis
- Preprocessed and analyzed fMRI image voxels to identify active regions of the participants' brains