Min Gu (Min) Jo

(510) 365-4988 | mingu08@berkeley.edu
http://mingujo.github.io | GitHub : github.com/mingujo

SUMMARY

Software/Data Engineer specialized in building data pipeline along with machine learning analytical process (mostly related to NLP), and web development.

EDUCATION

University of California, Berkeley **Class of 2016**

- ***B.A. in Computer Science, Statistics, and Economics***
- **Cumulative GPA: 3.42/4.00**

Relevant Coursework: Machine Learning • Software Engineering • Database • Algorithms & Data Structures • Statistical Inference & Computing • Linear Modeling • Web Development • Computer Network

SKILLS

Programming Languages

- Python, SQL, Java, R, Bash, Ruby

Frameworks, Libraries, Databases

- Apache Spark, Django, ElasticSearch, TensorFlow, Heroku, AWS(EMR, EC2, S3), Airflow, PostgreSQL, Amazon Redshift

PROFESSIONAL EXPERIENCE

June 2016 - Present
Berkeley, CA

LeadGenius
Data Engineer + Software Engineer Intern

- Implemented a web API which automates multi-classification of B2B sales outreach email responses by regularly trained RNN model with LSTM architecture (TensorFlow)
- Built a Back-End system which automatically notifies clients of new positive outreach responses
- Automated migration of labeled email responses from Nylas client to Postgres database for training
- Designed distributed ETL pipeline of U.S. based company and professional data from a variety of sources by applying cluster computing (Airflow, SparkSQL)
- Linked and merged duplicate companies using locality sensitive hashing on Minhash (Spark MLlib)
- Implement keyword and competitor search for U.S. based business using BoW and TFIDF
- Installed search engine on Django REST API to provide distributed search feature on the company database (Elasticsearch, Django-Haystack)

January 2016-December 2017
Berkeley, CA

Berkeley Institute of Data Science | OskiLab
Undergraduate Researcher

- Built binary sentimental classification model to label user reviews using Bag of Words model
- Implemented ETL pipeline of movie review data from web scraping to S3 (Scrapy, S3)

RESEARCH/PROJECT EXPERIENCE

Spring 2016

Kaggle Challenge: Rossmann Drugstore Store Sales Prediction

- Used 3 different machine learning algorithms to forecast drugstore daily sales: multivariate linear regression, Random Forest regression, and Gradient Boosting with regression trees

Fall 2015

Detection of Activated Brain Regions Under Mixed Gamble Task (In-class Project)

- 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

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

Probabilistic Modeling of Interactions on UC Berkeley Campus
Prof. David Aldous: Undergraduate Research Group

- Designed an independent research topic to predict and visualize common routes of UC Berkeley undergraduates with different majors and their interactions on campus
- Collected survey data from 130+ undergraduates across 5 different majors on MySQL database

Spring 2013

Prediction of Kobe Bryant's Performance in His Next Game (In-class Project)

- Scraped Kobe's seasonal data from *basketball-reference.com* and selected relevant predictors
- Applied regression analysis and feature shrinkage methods to create statistical models for prediction using R