# Min Gu (Min) Jo

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

## SUMMARY

Software/Data Engineer specialized in design, architecture, and development of data pipeline along with machine learning analytical process, and web application development

## EDUCATION

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

*Relevant Coursework*: Machine Learning • Software Engineering • Statistical Inference & Computing • Artificial Intelligence • Linear Modeling • Database • Computer Architecture • Computer Network

## SKILLS

**Programming Languages**
- Python, Java, R, SQL, Ruby, C

**Framework, Library, Database, Platform**
- Spark, Django, TensorFlow, ElasticSearch, AWS EMR, Airflow, PostgreSQL, Amazon Redshift

## PROFESSIONAL EXPERIENCE

*June 2016 - Present*
*Berkeley, CA*

**LeadGenius**
*Data Engineer + Software Engineer Intern*
- Implemented a web API which automates labeling of B2B sales outreach email responses using an RNN model with LSTM architecture using TensorFlow
- Built a Back-End system which automatically notifies clients of new positive outreach responses
- Enabled an automatic migration of all labeled email thread data from Nylas to Postgres database
- Predicted the open rate of outreach emails based on subject-line contents using multivariate linear regression
- Designed the parsing/merging ETL pipeline of U.S. based companies and professionals data from a variety of sources using locality sensitive hashing by scheduling parallel process
- Build the graphical database of U.S. companies for keyword clustering and recommender system.
- Managed Django REST API to provide distributed search feature on the company database

*January 2016-December 2017*
*Berkeley, CA*

**Berkeley Institute of Data Science | OskiLab**
*Undergraduate Researcher*
- Detected randomly positioned article texts in Bon Appetit magazines using Keras
- Build sentiment analysis model to classify user reviews using the Bag of Words model

## 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