Minjie Zhang

3355 The Alameda, Apt 105, Santa Clara, 95050
(714)234-1789 | vergil9312@yahoo.com

OBJECTIVE

Seeking a data scientist internship in 2017 Summer

EDUCATION

Santa Clara University – Santa Clara, CA — Expected Jun 2018
Master of Science in Computer Science & Engineering, Data Science — *GPA: In-progress*

University of California, Los Angeles – Los Angeles, CA — Sep 2012 – Aug 2016
Bachelor of Arts in Business Economics, Minor in Statistics — *GPA: 3.46*

Relevant Coursework: Algorithm, Machine Learning, Data Structures, Statistical Models and Data Mining, Data Analysis and Regression, Numerical Linear Algebra, Optimization, Natural Language Processing

TECHINICAL SKILLS

- Most experienced with **Python (pandas, scikit-learn, numpy)**, **R**, **C++**
- Some experience with **SQL**, **MatLab**
- Data analytics skills: **regression**, **classification**, **visualization**, **manipulation**

PROJECTS

Natural Language Processing Projects — Sep 2016 – Oct 2016

- Implemented a stemmer and **Viterbi algorithm** for computing the most likely parsing by probabilistic context free grammar and generated parse trees using NLTK and Python
- Implemented a **Text-Rank summarization** algorithm for news datasets using NLTK and Python

Kaggle Competition — Mar 2016 – Jun 2016

- Utilized 16 variables including shelter location, age, species, breed, color, year to predict whether an animal in shelter would be euthanized
- Applied **random forest**, **multinomial logistic regression** and **extreme gradient boosting** to 113,891 training data and predict results for 100,000 testing data
- Generated a multi-logloss of 0.31568, ranked 14th among 55 teams

LLC for Image Classification — Jan 2016 – Feb 2016

- Analyzed the paper *Locality-constrained Linear Coding for Image Classification* (2010) and implemented its encoding algorithm using MATLAB and Python
- Applied one-vs-all **support vector machine** classification method to data set Caltech-101
- Generated a 71% image classification accuracy for 101 categories with 5000 images

Coursera Machine Learning Projects — Dec 2015 – Jan 2016

- Applied one-vs-all **logistic regression** and **neural networks** to recognize handwritten digits
- Built an **anomaly detection algorithm** and applied it to detect failing servers on a network
- Used collaborative filtering to build a **recommendation system** with 1682 movies and 943 users
- Implemented **K-means** clustering algorithm and applied it to compress an image, and used **principal component analysis** to find a low-dimensional representation of face images