Cao Jilin

2315 Bowditch Street, Berkeley, CA 94704

caojilin@berkeley.edu caojilin 510-859-5316

EDUCATION

University of California, Berkeley **Berkeley, CA**
Bachelor of Arts in Statistics, GPA:3.47 *Sep 2017 - May 2019*

Santa Monica College **Los Angeles, CA**
Bachelor of Arts in Statistics, GPA:4.0 *Sep 2015 - May 2017*

Key Skills

Python, R, Java, Git, Hypothesis Testing, SQL, AWS, Mxnet, Pytorch

Related Coursework

Deep Learning, Machine Learning, Data Structure, Probability, Linear Algebra, Linear Programming and Network Flow, Linear Modeling, Efficient Algorithms and Intractable Problems

PROJECTS

Cloud Detection | R *April 2019 – May 2019*

- Used NASA satellite image data to develop a model to predict whether a given pixel is cloudy. A pixel in the image represents a 256m x 256m area in any place in the world. Fully investigated popular machine learning models such as KNN, SVM, Losgistic Regression, LDA/QDA, and Random Forests along with their performance and detailed diagnostics.
- Link: https://github.com/caojilin/cloud-detection

Predict Ames House Prices | Python *April 2018 - May 2018*

- Preprocessed the dataset by imputation, normalizing skewed predictors and factorizing categorical predictors. Used Ridge and Lasso regression with cross-validation to select the best model with the minimal RMSE and achieved 0.12 RMSE on the test dataset.

Image Captioning| Python *March 2019 – April 2018*

- Trained on the MSCOCO datasets with deep CNN and RNN models. The model can describe a given image with a detailed sentence. Detail in the github link.
- https://github.com/caojilin/image-captioning

EXPERIENCE

Data Analyst Intern **HangZhou, China**
State Grid Corporation *June 2017 - July 2017*

- Our task is to extract meaningful information from a large real world dataset. We performed unsupervised clustering methods like PCA, K-means and EM. Discovered many interesting findings from graphs we drew.

Lab Assistant for CS 61B and Stat 133 **Berkeley, CA**
Sep 2018 - Now

- Every week, I helped students understand the material covered in class and helped them work through their weekly labs and also worked at Office Hours. Stat133 is an introduction to computational data analysis class. cs61b is a data structure class.

AWARDS AND ACHIEVEMENTS

- 2018 Cal DataFest Team of Best Visualization
- Language: Fluent in Chinese