

Colin Lee

colin.lee.work@gmail.com | (530)277-9051 | colin-lee-13.github.io/pages/

OBJECTIVE

I am looking to secure a role that utilizes my data science skills developed through coursework and our capstone project, enabling opportunities to pursue a long-term career in data analytics.

EDUCATION

Master of Science in Data Science

Jan 2020

University of the Pacific – School of Engineering and Computer Science, Sacramento, CA

Bachelor of Arts in Applied Mathematics

May 2016

Minor in Applied Statistics

Humboldt State University, Arcata, CA

SKILLS

Machine Learning: Theoretical and practical understanding of supervised and unsupervised machine learning techniques such as Regression, Naïve Bayes, Random Forests, Decision Trees, Neural Networks, SVM's, Time Series, Clustering, etc.

Coding: Python, R, SQL

Data Visualization: Tableau, R-Studio

Database Software: MySQL

Web Scraping

Statistical Significance Testing

Jupyter Notebooks

EXPERIENCE

Sacramento Kings Capstone Project

Sept 2019-Dec 2019

Objective: Predict ticker-holder attendance

Received ticket sales and email marketing data from past two seasons

- Cleaned data by formatting data types, removing incomplete entries
- Identified and removed outlier instances through EDA and based on instruction

Conducted custom feature engineering

- Tried to think all potential factors that a ticketholder might consider that could affect their decision on whether or not to attend a given game
- In addition to ticket sales and marketing features that were provided, we engineered demographic, weather, seating, and on-court basketball related features in an attempt to capture the complete thought process of the ticketholder
- Data set was not available for basketball related features, so we engineered a web scrape of basketballreference.com in order to get the data we needed for basketball related features

Modeled ticket-holder attendance using multiple machine learning techniques

- We attempted to model ticket holder attendance by utilizing random forests, neural networks, naïve bayes classifiers, SVM's, and logistic regression
- Our data was naturally unbalanced so we also used three different sampling methods for all of our model approaches

Presented and communicated entire process and results

- Presented PowerPoint explaining thought process, actions, and results to Business Ops Team
- Achieved ~92% accuracy of predictions with our best performing model
- Entire project was written using Python in Jupyter Notebooks and use of Tableau for visualizations

Greenskeeper

May 2018-present

Tahoe Mountain Club, Truckee, CA

Sales Associate

Nov 2016-Aug 2017

Big 5 Sporting Goods, Grass Valley, CA