Michelle K. Li

DATA SCIENTIST

michelle.li.524@gmail.com
626-888-9862
Monterey Park, CA

github.com/michellekli
linkedin.com/in/michelleli524
michelleli.tech

Data scientist candidate skilled in implementing machine learning to solve business problems. Experienced with Python, SQL, scikit-learn, seaborn, matplotlib, and experimental design.

Skills

Data Science: statistics, experimental design, data wrangling, exploratory data analysis, data visualization, presenting results
Machine Learning: supervised/unsupervised, regression, classification, clustering, natural language processing, time series analysis
Python Packages: numpy, pandas, matplotlib, seaborn, scikit-learn, statsmodels, SciPy, spaCy
Programming: Python, SQL, Java, C/C++, JavaScript, HTML/CSS, Git, OOP, API design, ETL scripting

Recent Projects

***Time Series Analysis and Forecasting for Restaurants*:** (github.com/michellekli/visitor-forecasting)

- Compared time series models (ARIMA, SARIMAX, BSTS) for forecasting daily visitors at restaurants
- Key finding: seasonal models (SARIMAX, BSTS) performed better than ARIMA
- Created pipeline to generate forecasts for any of the 800+ restaurants in the dataset

***Text Classification on Novels*:** (github.com/michellekli/love-stories)

- Compared 7 models (logistic regression, multinomial/Bernoulli/Gaussian Naive Bayes, SVM, random forest, MLP) against 5 clustering algorithms (mean shift, spectral clustering, K-means, affinity propagation, DBSCAN) for classifying novel text by author
- Key finding: multinomial Naive Bayes performed 40% better than the best spectral clustering algorithm
- Extracted features with NLP approaches: bag of words, tf-idf, word2vec, positive PMI
- Adapted techniques from research papers to estimate accuracy of clustering algorithms

***Linear Inference on House Prices*:** (github.com/michellekli/melbourne-housing)

- Compared OLS and ridge regression for inferring effect of features on house prices
- Key finding: 1 m^2 more building area = ~0.5% higher price; 1 *km* further from suburb = ~0.75% lower price
- Created ridge regression model explaining 69.5% of the variance in house prices

Experience

CGI Inc. — Los Angeles, CA
Software Engineer & Technical Consultant — 07/2016 – 01/2019

- Designed and implemented audit architecture for data analytics across mobile/backend/middleware layers
- Wrote complex SQL queries to identify low quality data while fixing tens of bugs
- Partnered with external developers from BAVN to design project specifications for long-term ETL process
- Interfaced with BAVN's external API to perform ETL for batch data processing every 10 minutes
- Scaled mobile application for concurrent usage across 10 locations by identifying and removing performance bottleneck
- Key team member responsible for presenting more than 5 client-facing demos

Education

University of California, Los Angeles (UCLA)
BSc in Computer Science — 2016
cum laude (GPA: 3.7 / 4.0)

Thinkful — 2019
Data Science Program