Jaspreet Kang

Data Analytics · Data Science
San Francisco Area
(925) 381-0646 | jaspreetkang@ucla.edu | kangjaspreet | jaspreet-kang1

Education

University of California, Los Angeles — *Los Angeles, CA*
Bachelor of Science in Statistics — *Graduated: June 2019*
- GPA: 3.20

Work Experience

Logos News, LLC — *Los Angeles, CA*
Data Science Consultant — *April 2019 - Present*
- Extracted and processed over one million different types of data (numeric, textual) in Python from several news media sources by utilizing developer APIs.
- Built a Comment-Ranking Model, a model that quantifies the quality of a comment by predicting the: toxicity, subjectivity, political affiliation, and quality of a comment.
  - Utilized Natural Language Processing packages in Python (Keras, Tensorflow) to pre-process textual data and transform words into vectors using Word2Vec.
  - Yielded over 90% successful classification rate by applying a neural network algorithm called Long Short-Term Memory model to training and testing data.
- Led the development of the User Reputation Model project, a project aimed to increase consumer experience and satisfaction on the app by calculating a user's reputation score for app personalization purposes.
  - Employed project management skills to coordinate project tasks. Researched appropriate methods to gather information in order to develop project scope, and conduct work plans. Provided status updates to management frequently.

Ornus Capital, LLC — *Los Angeles, CA*
Data Science Intern — *April 2019 - June 2019*
- Found and optimized a portfolio of assets, which outperformed the market by over 30%.
  - Applied portfolio theory (e.g. Markowitz theory), regression analysis, and bootstrap/simulation methods in order to find portfolios with highest expected returns, and portfolios with lowest risks (minimum variance).
- Automated a program in Python that extracts, translates, and loads historical data of user-inputted cryptocurrencies within a specific time period.
  - This program automatically outputs a heatmap of the correlation coefficients, time-series plots, and scatterplots with fitted OLS regression lines in order to get a better understanding of asset and variable relationships, and to identify trends and correlations.

Clayton Treat Automotives — *Concord, CA*
Business Analyst Intern — *June 2018 - September 2018*
- Increased labor work and revenue by over 10% by conducting local market research and statistical analysis.
  - Performed time-series analysis in Microsoft Excel to determine which types of jobs combined with time periods generated most revenue.
  - Suggested and implemented recommendations, which included: creating promotional offers during certain time periods with lagging sales, changing price tags of many car parts and labor costs, implementing workflow solutions and process improvement opportunities, and more.
- Assisted in the creation of a business website and placed promotional offers on there.

Projects

Metro Bike Share Consulting

- Consulted the company Metro Bike Share directly and provided recommendations to increase business performance such as: optimizing their spatial allocation of bikes based on the customers' activity time periods, refining memberships costs and perks in order to increase revenue from both bike rides and subscriptions, and utilizing promotional offers during stagnant seasons.
- Applied clustering and time-series analysis on over 650,000 data points. Aggregated data to find key performance indicators and trends within each region of operation.

Housing Affordability in Seattle (Decision Trees in R)

- Classified housing affordability correctly with an average of around 98% by applying a refined Random Forest model and performing k-fold cross validation. Narrowed down to the most impactful predictive variables by creating a Variable Importance Plot.
- Cleaned up and manipulated over thousands of rows of data by performing extensive data cleaning using Mice, Amelia, and Miss Forest packages in R. Process included handling over 25,000 missing values.

Skills

Machine Learning Linear Regression, Logistic Regression, Random Forest, K-Means Clustering, Regularization, Neural Networks, NLP
Programming R (dplyer, ggplot2), Python (Pandas, Numpy, BeautifulSoup, TensorFlow, matplotlib, seaborn), SQL, SAS, HTML, CSS
Other Skills Microsoft Excel, Microsoft Word, PowerPoint, Tableau