SHERRY DUONG

408-442-8444 | sherryduong@gmail | sherryduong | sherryduong93

QUALIFICATIONS

Data Science Applications:

- Analytics: Data Visualization, Statistical Inference, Hypothesis Testing, Sampling Methods, A/B Testing
- Machine Learning: Linear/Logistic Regression/Classification, Regularization: Ridge & Lasso, Decision Trees, Random Forest, Neural Networks (RNN/CNN), Natural Language Processing, Time Series, Recommender Systems, Graphs

Programming & Analytics:

- Python Libraries: NumPy, Pandas, Scikit-Learn, Tensor-Flow, NLTK, Matplotlib, Seaborn, Beautiful Soup, Flask
- Database Management: Postgres & Psycopg2, Microsoft SQL Server, Spark & PySpark, MongoDB & Pymongo
- Additional Technologies: Tableau, Amazon AWS (Sagemaker, EC2, S3), Docker, Heroku, Netsuite

PROJECTS

Analyzing the Chase Center Impact on Crime in Dogpatch/Mission Bay -

- Conducted hypothesis testing with python's Scipy library, using Welch's T-test and Mann Whitney U-test to confirm whether or not the opening of the Chase Stadium had a statistically significant impact on crime rates in the Dogpatch/Mission Bay areas. Confirmed that while events are seemingly disruptive, they are not as impactful.

Predicting AirBnB Listing Prices in San Francisco - |

- Analyzed current Airbnb daily listing prices in San Francisco to predict prices of future listings using SKlearn's Linear Regression, Random Forest, and Gradient Boosting. Performed feature engineering and hyper-parameter tuning, eventually achieving 75% improvement on RMSE for final RMSE of $54 on cross-validated data.
- Additional Tools/Libraries Used: NLTK, Tensorflow

Your Anime Match Maker: An Anime Recommender - | |

- Created a recommender system to connect anime lovers to new anime based on popularity, content using cosine similarity, and matrix factorization collaborative filtering. Used Spark's ALS collaborative filter model and achieved 1.13 RMSE on predictions of 1.3M user ratings (from 1-10). Web app created with Flask and is live on Heroku.
- Additional Tools/Libraries Used: Spark ML, AWS Sagemaker & S3, Flask, Heroku

EXPERIENCE

Galvanize Inc. **2020 - Present**

Data Science Resident

- Selected among 25 other data science students based on academic achievement and understanding of concepts.
- Instructing data science students on topics including python programming, machine learning, and statistical inference.
- Innovating on curriculum material, providing targeted content in order to reinforce retention of program materials.

Data Science Fellow

- Projected profit of $200K for a ticket sales and distribution company by identifying fraudulent users using machine learning and cost-benefit matrix. Achieving an AUC of 99.4% with random forest classifier.
- Classified sentiments of movie reviews using topic modeling, clustering, and classification models.
- Predicted user churn for ridesharing company using gradient boosting classification model.
- Forecasted HIV incidence in US counties using Linear Regression (sklearn & statsmodel).

IPSY (Personalized Beauty Discovery Inc.) **2017 - 2020**

Project Manager (2018-2020) | Procurement/Logistics Analyst (2017-2018)

- Analyzed shipment data to optimize quality control. Identified a single quality issue responsible for 60% of all costs, implemented changes that reduced cost by 40% in one year, over $150K in costs annually.
- Built monthly reports using Microstrategy and Excel to streamline inventory movements for the subscription program, resulting in a significant reduction of errors by 30% for over 6M units of inventory monthly.
- Optimized supply chain operations by examining pricing and identifying risks/bottlenecks in the supply chain process among 200+ SKUs, 20+ global suppliers, and 20+ internal team members.

EDUCATION

University of California: Berkeley: Bachelor's of Science: Business Analytics **2011 - 2015**