Donald Vu

Los Angeles CA 90024 • (714) 658-4599 • donaldvu18@gmail.com

Education

Georgia Institute of Technology **Present - June 2020**
Candidate for Master of Science, Computer Science
Concentration: Computing Systems

University of California, Los Angeles **September 2013- June 2018**
Bachelor of Science, Statistics

Experience

The Walt Disney Company **May 2019 - August 2019**

***Data Science Intern**; Burbank, California*

- Oversaw and maintained complete ownership over the production of a statistical model from idea formulation to pipeline integration and full deployment.
- Constructed Net Uplift and Pareto/NBD models to predict customer lifetime value, cost per registered user, and customer retention/engagement in order to measure the most effective channels for Disney's marketing media.
- Developed fluency with AWS EC2/S3 services by deploying large-scale models on high-level instances, mining volumes of raw data for feature engineering, and executing vast quantities of database queries.

RAPP – Advanced Analytics **August 2018 – May 2019**

***Data Science Intern**; Los Angeles, California*

- Coded analytical models in Python for McDonalds, Toyota, and Nestle that enhanced their customer targeting precision and solved business challenges/KPIs through the use of LightGBM, XGBoost, and clustering algorithms.
- Implemented multi-core processing and memory manipulation for in-house scripts reducing average code runtime from 2 hours to less than 30 minutes while also conserving 30% more data storage.
- Utilized SQL to execute batch queries on Hadoop/Spark and extracted 40 million+ records for use in Python pipelines to support feature enhancement of existing models.

Projects

Portfolio: https://donaldvu18.wixsite.com/mysite

Twitter Sentiment Analysis Project using NLP/NLTK/Scikit/SciPy/Pandas

- Analyzed community outlook on various topics by using Python to capture tweets using Twitter's streaming API and then training a classifier to recognize their connotation.

NBA Playoff Wins Algorithm (www.bruinsportsanalytics.com/defchamp **)**

- Developed a regression model that predicts the number of wins for a playoff team based on multivariate predictors and had the algorithm published on a sports analytics website.

Skills

Python (4yrs), R (3yrs), SQL (3yrs) | AWS / EC2 | Hive/Hadoop | PySpark | Decision Tree
SVM | Dashboards | TensorFlow | Tableau | K-NN | Logistic Regression| NLP | Neural Networks