DILLON QUAN

dquan3@dons.usfca.edu | 510-507-3299 | Oakland, CA | www.linkedin.com/in/dillon-quan/

EDUCATION

University of San Francisco Expected Jun.2020
M.S Data Science
Related Coursework: Machine Learning, Deep Learning, SQL, Distributed Computing (PySpark), Experimental Design (A/B Testing), Data Structures & Algorithms, Data Visualization, Natural Language Processing (NLP)

University of California, San Diego Jun.2015
B.S Electrical Engineering

EXPERIENCE

TruSTAR Technology, *Data Science Intern*, San Francisco, CA Nov. 2019 - Current
- Currently solving the malicious URL detection problem by performing various feature engineering techniques with Machine learning models like Logistic Regression and Random Forest to improve the solution's interpretability.
- Built a 1-D character-level Convolutional Neural Network (CNN) model to help security analysts detect malicious URLs with 95% accuracy using purely lexical features using Pytorch.
- Developed parsers to extract and normalize raw JSON data to process into a datalake that will help security analysts sift through cyber threats using Spark.

E-HealthNow, *Data Science Intern*, Oakland, CA Aug. 2018 - Feb. 2019
- Trained a Transformer NMT model using OpenNMT framework in Pytorch. Achieved a BLEU score of 25 using the AI challenger dataset.
- Contributed to the product roadmap development process with the management team to devise a pipeline to transfer patients' medical records to doctors.

Knolls Atomic Power Laboratory, *Electrical Engineer*, Schenectady, NY Oct. 2016 - Mar. 2018
- Analyzed and verified measured data are within specified ranges using Excel.
- Developed procedural plans and performed verification testing to determine the compatibility of overall systems and troubleshoot systems.
- Supported the Ohio Class Submarines and the fleet by troubleshooting systems and performing failure analysis.

PROJECTS

Streamhopper Mar.2020 -May. 2020
- Successfully developed and pitched a minimal viable product to venture capitalists that would recommend users specific streaming service each month to solve the subscription fatigue problem.
- Implemented a content-based recommendation system using TFIDF-scores on the plot for each content.
- Generated a questionnaire on the website to turn user data into personas using k-modes clustering.

Scikit-Learn Clone Oct. 2019 - Feb. 2020
- Implemented from scratch various supervised learning and unsupervised learning algorithms with comparable performance to scikit-learn such as regularized Linear Regression, Logistic Regression, Naive Bayes, Decision Trees, Random Forest, and kmeans clustering using numpy.
- Coded drop column and permutation feature importance using pandas to compare and contrast pros and cons.

Predicting the Next NBA MVP Feb. 2018 - Apr. 2018
- Used Logistic Regression to successfully predict James Harden to be the 2018 NBA MVP before the award was given.
- Webscraped Basketball-Reference.com using BeautifulSoup to data-mine players stat line for feature engineering.

SKILLS

TOOLS: Python (scikit-learn, numpy, pandas, matplotlib), Pytorch, Pyspark, SQL (Postgres), ggplot2
OTHERS: Amazon Web Services (AWS), Databricks, Web-Scrapping (Selenium, BeautifulSoup, requests)