Justin Gong

Data Science | Advanced Analytics | Machine Learning

Data scientist with 2 years experience in project work. Skilled in finance, machine learning, statistics, and programming. Strong curiosity to learn, highly coachable, and excited to bring value to your team.

gong-jj.com · jjgong7@gmail.com · (510) 517-0077 · New York City · linkedin.com/in/justin-gong · github.com/jjgong7

EDUCATION

Master of Science in Analytics   *GPA 3.83*
Georgia Institute of Technology   *05/2019*

Bachelor of Science in Managerial Economics   *GPA 3.54*
University of California, Davis   *06/2014*

WORK EXPERIENCE

Data Scientist

Kurvv (Contract) | 11/2019 – Present | Remote
Developed script for e-commerce clients to analyze customer purchasing history to upsell and cross sell products.
- Implemented market basket analysis and collaborative filtering (ALS) on implicit order/purchasing data to make personalized product recommendations for customers using Python.
- Improved accuracy of product recommendations for a client by ~5% using CF versus recommending popular items.

Associate Accountant (CA Certified Public Accountant License: #130384)

Devereaux, Kuhner and Co. LLP | 07/2014 – 02/2017 | San Francisco, CA
- Reviewed financial statements and data of various companies. Prepared federal and state income tax returns.

PROJECTS

Emory Healthcare Practicum (01/2019 – 05/2019) | Link: www.gong-jj.com/nhanes

Cleaned and processed 1,222 files (3.2 GB) of NHANES healthcare data across 18 years into MongoDB. Identified the risk factors and prevalence of hospital utilization and four major diseases over time using qualitative and quantitative data.
- Implemented Random Forest and XGBoost in Scikit-learn to predict hospital utilization with an f1-score of 0.52.
- Visualized increasing prevalence of cancer and diabetes and constant prevalence of heart disease and chronic lung disease over time using ggplot2 and survey libraries in R.
- Identified age, gender, and blood pressure as common risk factors across the four major diseases.

Supervised Learning (09/2018) | Link: www.gong-jj.com/sl/

Analyzed Faulty Steel Plates and Phishing Websites datasets. Implemented Decision Trees, Boosting, Neural Networks, Support Vector Machines, and k-Nearest Neighbors using Weka and compared results.
- Evaluated bias and variance tradeoff (learning curve) and performed hyperparameter tuning (model complexity).
- Achieved 77.12% accuracy using Boosting with Decision Trees for Faulty Plates, an improvement of 7.97% versus SVM.
- Achieved 97.06% accuracy using k-NN for Phishing Websites, an improvement of 2.76% versus SVM.

M.O.R.E: An Application for Career Path Exploration (02/2018 – 04/2018) | Link: http://bit.ly/more-appl

Collaborated with team to design an interactive college major and career exploration and recommendation application. Cleaned and processed more than 15 million records and 280 columns from the 2012-2016 ACS 5-year PUMS data.
- Improved existing applications by using interactive exploration, more statistics, and similarity algorithm.
- Succeeded in changing 34% of participants' answers related to best major-career paths, 41% of participants' answers related to earnings for a major, and resulted in ~62% of users feeling more informed.
- Implemented front-end using jQuery, D3.js, and Bootstrap and back-end using a Python Flask app with a SQLite3 DB.

SKILLS

Machine Learning: Regression, Classification, Clustering, Decision Trees, Random Forest, SVM, k-NN, Neural Networks
Software and Programming Languages: Python (Scikit-learn, SciPy, NumPy, Pandas, Matplotlib), R (dplyr, ggplot2), SQL (SQLite, MySQL), NoSQL (MongoDB), JavaScript (D3), Git, Jupyter Notebooks, Tableau, AWS, Weka, Excel