

SKILLS

Python, R, SQL, Machine Learning, Jupyter Notebooks, PyCharm, R Studio, Flask, scikit-learn, statsmodels, transformers, tensorflow, pyTorch, Google Cloud, AWS Sagemaker, Tableau, Github

EDUCATION

San José State University Aug. 2018 to May 2020

Masters Statistics 2020

Courses Taken: Machine Learning(Stanford), Cluster Analysis, Time Series, Categorical Analysis, Computational Statistics, Design and Analysis of Experiments, Regression Theory and Methods

GPA: 4.0

De Anza College Jan. 2016 to Mar. 2018

Courses Taken: Python Programming, Intermediate Programming Methodologies in C++, Data Abstraction and Structures, Operating System Concepts, etc

GPA: 3.93

EXPERIENCE

METIS

Data Scientist Fellow

San Francisco, CA
June 2020 to Sept. 2020

- Completed a 12-week immersive ACCET-accredited data science program.
- Designed and built practical business solutions for e-commerce platforms.

INTUIT

Contractor

Mountain View, CA
Aug. 2019 to Dec. 2019

- Optimized loss function of a novel algorithm developed by senior data scientists at Intuit using PyTorch from scratch.
- Demonstrated the time complexity of the algorithm from the theory in code with Boston Housing Price dataset.

SAN JOSÉ STATE UNIVERSITY

Teaching Assistant

San José, CA
Aug. 2018 to May 2019

- Gained communication skills by teaching important concepts in Calculus to business school students and helped them pass.

BITNINE

Researcher/Web Developer

Seoul, South Korea
Oct. 2014 to July 2015

- Exclusively developed a PostgreSQL database manager with HTML, javascript, CSS from scratch.
- Operated Ubuntu terminal to execute queries in backend.
- Presented user-cases of graph database to executives with effective visualizations.

PROJECTS

REVIEW-BASED SEARCH ENGINE

Sept. 2020

- Developed a search engine that outputs products that match the key qualities with positive sentiment from the reviews.
- Created a user-interactive Flask app that retrieves optimal products with summarized reviews
- Tools: Universal Sentence Encoding(USE) for sentiment analysis, PEGASUS for summarization, Spacy and NLTK for preprocessing

NETFLIX APP REVIEW TOPIC MODELING

Aug. 2020

- Motivated to help developers to promptly deal with trouble-shooting and adjust their quality assurance.
- Trained BERT model to classify the reviews with fine-tuned LDA topic labels.
- Created an app for users to write a review and be referred to the right service. Wrote a blog about the difference between the bag-of-words scheme and context-base scheme. Published on Chatbot's Life Magazine.

IEEE-CIS FRAUD DETECTION ON KAGGLE

July 2020

- Using online-transaction dataset on Kaggle, compared Bagging methods with respect to recall and precision. Achieved 0.87 AUC with XGBoost and Sigmoid Calibration on Kaggle submission. Permutation importance with recall ratio was used for interpreting each feature's contribution.
- Tools: xgboost, eli5, sklearn.calibration

KEYBOARD SALE RATE PREDICTION BY POISSON REGRESSION

July 2020

- Predicted a keyboard's sale rate using Poisson regression. Handled feature engineering, imputation, and cross-validation.
- Web-scraped eBay query results using BeautifulSoup and improved R^2 and MAE by feature engineering and multiple imputation with MICE algorithm. Selected by Medium curators to be recommended to readers in 'Data Science' section.