Junlan Lu (Jaqueline)

Email: jellalu21@gmail.com Phone: (626)-715-3847 URL: https://jaqueline-lu.github.io

Education

University of California, Los Angeles as *Computer Science Major* **(**GPA: 3.3) Fall 2018 - Present

Experience

Data Specialist at *Rainforest Connection (RFCx)* March 2020 - Present

Train the AI model (chainsaw), review alerts from the acoustic monitoring machines put in the ecosystems to improve the ML models and create more new ML models. Perform data cleaning, data selection, feature engineering and model selection on acoustic data collected. Work with other engineers, developers and hackers to tackle problems from the machines and models.

Projects

Categorical Feature Encoding Challenge (Winter 2020)

Using Tensorflow backend, I build a deep learning model for this datasets. Since there are already too many columns, I do label encode for the features, and use Kfold validation to evaluate the models. As a comparison, I also perform one hot encode (with more focus on the features themselves) and run catboost model on it. After tuning hyper-parameters, it also does a good job in prediction.

NYC Taxi Rides Duration Prediction (Winter 2020)

The large dataset includes about 100k rows. I perform data selection and EDA based on data visiualization drawn to filter out the data that are most related, and do feature engineering (PCA to divide up the map of Manhattan into regions); and then I perform model selection on different regression models (linear, Lasso, Ridge, SVR etc.) and make better use of the features by putting different weights.

Kaggle: Titanic Disaster (Fall 2019)

Python with scikit-learn to do the machine learning model, I choose Random Forest Classifer model, and perform feature engineering both to create effective features and to clean the noisy data. I also use pandas, matplotlib and seaborn modules to do data visualization. This model gives me the performance of 80.38 % as top 8% on Leaderboard in predicting the survival of the passengers.

Scrapy: House on Rent and on Sale (Summer 2019)

Using Python (requests and beautiful soup) and SQL to capture the housing rent and sale on certain websites, given the input from users specifying the name of the city and number of outputs. It collects the information and transforms to an Excel spreadsheet that is used to create data visualization.

Skills

- Languages: C/C++, Python (scikit-learn, matplotlib, pandas, seaborn), shell script, javascript
- Software: Microsoft Office, Adobe Premiere Pro

Related Courses

Data Science Fundamental, Intro to CS (C++ and data structure), Machine Learning (Coursera), Intro to Probability and Statistical Reasoning, Algorithm and Complexity