Nuoya(Noah) Rezsonya

Phone: 904-521-5938 I **Email:** kangnuoya@gmail.com I **Location:** San Jose, CA 95112
LinkedIn: https://www.linkedin.com/in/nuoya-noah-rezsonya/

Education

Southern Methodist University — **Dallas, Texas**
Master of Science in Data Science I Current GPA 4.0 — **Class of 2019**

University of New South Wales — **Sydney, Australia**
Bachelor of Civil Engineering I GPA 3.4 — **Class of 2014**

Experience

Data Analytics Intern — **San Jose, CA**
Visual Threat — **12/2018 - Present**

- Developed an analytical report of vehicle ECU stability.
- Increased the amount of effective data by refining the data collection strategy.
- Transformed vehicle-generated data into readable hex traffic data by writing Python script.
- Identified data changes and recorded changing patterns by comparing processed data with corresponding brand new vehicle with Meld on Linux.

Data Science Intern — **San Francisco, CA**
Ambit Analytics — **09/2018 -12/2018**

- Evaluated NLP model performance for speaker identification and identified factors related to overall model performance.
- Labeled audio data which were used as the ground truth for the performance evaluation by using Audacity.
- Tested model performance and visualized the difference between labeled data and prediction output with Python scikit-learn classification metric and Python matplotlib.
- Found 3 major factors related to overall model performance including background noise level, voice similarity and random interruption during a conversation. Provided solutions to address each of issues.
- Increased the accuracy of the model by 15.87% by implementing the proposed solution of random interruption.

Post-Graduate Structural Intern — **Chongqing, China**
JizhunFangZhong Engineering and Architecture — **10/2015 - 08/2016**

- Built a database of steel design parameters and costing by using MySQL.
- Supported development and management of scope, schedule and budgets for projects.
- Monitored construction progress against procurement schedule.

Projects (Github: https://github.com/nuoyakang/EducationProject)

Improving Teacher Recruitment and Retention in North Carolina Public Schools — **12/2018 - Present**

- Investigated factors that influence teacher recruitment and retention in NC public high schools by using NC public high school data, local IRS data and criminal rate data.
- Merged, cleaned data from 2014 to 2018 and created data dictionary by using Python pandas.
- Selected 80 features from school funding, school environment and performance, school expenditure on salary, overall teacher qualification and highest degree, minority division, tax paid by district and criminal rate by district for Logistic Regression model by using recursive feature elimination method with Python scikit-learn.
- Established the Logistic Regression model with stratified shuffle split and visualized feature importance by using Python scikit-learn and matplotlib.
- Evaluated the model performance by using Python scikit-learn evaluation metric.

Identifying factors influencing school performance of North Carolina Public Schools **04/2018 - 08/2018**

- Investigated factors that influence school performance of NC public high schools by using NC high school data.
- Merged and cleaned data from 2014 to 2017 by using Python pandas.
- Selected 75 features including school funding, school environment and performance, overall teacher qualification and minority division for Logistic Regression model and Random Forest model by using recursive feature elimination method and manually selected features with Python scikit-learn.
- Established models with shuffle split and visualized feature importance by using Python scikit-learn and matplotlib.
- Evaluated the model performance by using Python scikit-learn evaluation metric.

Skills

Analytic Tools	**Python**(pandas, numpy, scikit-learn), **R(**ggplot2, plyr, forecast, olsrr, caret), **SAS**
RDBMS	**MySQL**
Data Visualization	**Tableau, Python**(matplotlib), **Processing**
Machine Learning	**Regression:** Linear Regression, ANOVA, MANOVA **Classification:** Logistic Regression, SVM, k-NN, Random Forests, Ensemble Methods **Clustering:** k-Means, DBSCAN