

Noah Rezsonya

Project Links: <https://github.com/nuoyakang>

LinkedIn: <https://www.linkedin.com/in/noah-rezsonya-717447145/>

Legal Status: Greencard Holder

Location: San Jose, CA 95112

Phone: 904-521-5938

Email: kangnuoya@gmail.com

Skills

Programming Language

Python, R, MATLAB, Java

Analytic Packages

Python(pandas, numpy, scikit-learn, PySpark), R(ggplot2, rvest, plyr, forecast, olsrr, caret)

Database

MySQL workbench, MongoDB

Data Vizualization

Tableau, Python(seaborn, matplotlib)

Machine Learning

Regression: Linear Regression, ANOVA, MANOVA

Classification: Logistic Regression, SVMs, k-NN, Random Forests, Ensemble Methods

Clustering: k-Means, DBSCAN, Agglomerative Clustering

Education

08/2017 - 08/2019

Southern Methodist University: Dallas, Texas USA

Master of Science in Data Science | Current GPA 4.0

07/2009 - 02/2014

University of New South Wales: Sydney, Australia

Bachelor of Civil Engineering

Experience

09/2018- Present

Data Science Intern: San Francisco, CA USA

Ambit Analytics Inc.

- Annotating voice data.
- Curating a representative audio data set with ground truth information for evaluating the accuracy of different speech metrics.
- Identify gaps in accuracy from spec and tune parameters in the engine to improve accuracy.
- Enhance existing tools / build new tools to evaluate the data set against our speech engines.

Primary tools: Python(pandas, numpy, scikit-learn, matplotlib)

10/2015 - 08/2016

Post-Graduate Structural Intern : Chongqing, China

JizhunFangZhong Engineering and Architecture

- Maintain database of steel design parameters and costing
- Budget tracking, procurement and construction sequencing

Primary tools: Microsoft Excel, MySQL workbench

05/2014 - 11/2014

Executive Assistant Intern : Sydney, Australia

Harman Global Holdings

- Development of a new client database
- Document transactions and record feedback from customer interactions.

Primary tools: Microsoft Excel, MySQL workbench

Projects

05/2018 - 09/2018

Data Mining Using NC High School Data

Leveraging North Carolina high school data to investigate factors that influence high school performance.

Skills Used: Python scikit-learn, Data Pipeline, Feature Selection, Regression(Linear/Ridge/MLP Regressor), Classification(k-NN/Logistic/Random Forest), Clustering(as pre-processing step)

02/2018 - 05/2018

Mining Twitter Data With Python And MongoDB

Streaming tweets using Twitter REST API.

Compare design considerations: Real-time vs. Batch.

Skills Used: Python, MySQL workbench, MongoDB

02/2018 - 05/2018

Classify Kobe Shots

Using 20 years of basketball data from Kobe to classify shots.

Skills Used: SAS, Statistical Imputation, Feature Selection, PCA, Logistic Regression, LDA