

FRANK XU

✉ frankxu0124@gmail.com 🌐 medium.com/@frankxu0124 ☎ 9193083187 📍 Durham, NC (willing to relocate) in frank-xu-huaze/ 🌱 Frank-Xu-Huaze

SUMMARY

As a Master of Data Science at Duke, I have various experience cleaning and modeling data with Python & R: I ranked top 10% tier in a Kaggle Competition using Convolutional Neural Networks for object detection. I built machine learning pipelines with Google Cloud Platform and brought at least 5% potential profit increase for local companies. I know how to communicate to different audience from my experience working as Data Scientist for companies as big as Wells Fargo, as small as a local start-up, and EPAs. Need sponsorship in the future.

EXPERIENCE

Trade Pending, Data Science Intern, Carrboro, NC Aug. 2019 - Apr. 2020

Handled the 72-million rows dataset and deployed dashboards used Google Cloud Services, built a Random Forest and provided 92% accuracy on predicting vehicle transaction prices, bringing at least 5% potential profit increase for the product.

Wells Fargo, Data Science Intern, San Francisco, CA June 2019 - Aug. 2019

Analyzed customer data using Random Forests and Gradient Boosting Machines to recommend banking products. Researched on different methods of interpreting ensemble tree models, bringing more accurate decisions and potential profits for the team.

EPA of Dezhou City, Data Analyst Assistant, Shandong, China Jan. 2018 - Feb. 2018

Analyzed hourly air pollution with R and presented more than 20 charts every day for the government's project of improving air quality.

EDUCATION

Duke University 2018 - 2020

Master in Interdisciplinary Data Science

Enrolled with an acceptance rate less than 5%

Fudan University 2014 - 2018

Bachelor in Environmental Science

Outstanding Student Scholarship (2015-2017)

Shanghai University of Finance and Economics 2015 - 2017

Minor in Accountancy

PROJECTS

Exploring methods of Feature Interpretation Feb. 2020 - Mar. 2020

An article exploring new methods of feature interpretation for ensemble tree models available on Python and R.

Kaggle Competition: Identifying Solar PV in Aerial Imagery Jan. 2019 - Feb. 2019

Used a Convolutional Neural Network built based on MobileNet to achieve a 97.8% accuracy on identifying solar panels in aerial images and ranked 10% tier.

Analyzing Text Data for Real Talk Sept. 2018 - Feb. 2019

Real Talk is a mobile app available on iOS facing teenager users.

Ran text & regression analysis with R on a dataset of around 1400 stories, unveiled the connection between teenager psychometric status with demographic information.

SKILLS

Python, R, Tableau, Excel, Keras, TensorFlow, MySQL, SQLite, Hadoop, Spark