

Hai Huang

814-954-9069

Hoboken, NJ

hhuang37@stevens.edu

EDUCATION

Stevens Institute of Technology, Hoboken, NJ

Master of Science in Data Science

Expected May 2022

Related Courses: Statistical Methods, Probability Theory

GPA: 3.89

Pennsylvania State University, State College, PA

Bachelor of Science in Data Sciences

May 2020

Bachelor of Science in Applied Statistics

May 2020

Related Courses: Machine Learning, Data Visualization & Analytics, Data Integration & Fusion

SKILLS

Languages & Aptitudes: Python, R (R Studio), SQL, Hadoop, Scala, Tableau, HTML, CSS

Technical Skills: Regression (Linear, Multiple-Linear, Logistic, Random Forest)

Classification (Random Forest, Naïve Bayes, KNN), Data Cleaning & Interpretation

EXPERIENCE

Agriculture Bank of China, Guangdong Science & Technology Park

May 2019-August 2019

Data Scientist Intern

- Discussed and planned for a recommendation system for wealth management products
- Employed **LightGBM** model and used **Gridsearch** to tune parameters in Python for an internal machine learning competition
- Placed No.6 in the competition (predict how likely customers will purchase products)

ACADEMIC PROJECTS

Pennsylvania State University

Capstone Project: Estimate the Upper Bound of US Innovation

Spring 2020

- Collected and cleaned data from multiple datasets and conducted EDA on datasets
- Analyzed data trends based on region, income, education level, number of patents, etc.
- Identified the flaws of given data and convinced sponsor to change and find new datasets

PUBG (Player Unknown's Battlegrounds) Placement Prediction

Fall 2019

- Applied XGBoost, LightGBM and CatBoost algorithms in Python to predict final placement
- Implemented **ensemble learning** method with algorithms above to improve accuracy
- Developed and published a website for project by using HTML & CSS

PUBG Match Stats Datasets

Fall 2019

- Built interactive plots using Pyplot package in Python and employed plots in webpage
- Analyzed of the relationship between match data and death data through plots
- Enlisted and mentored struggling teammates

PROJECTS

Kaggle Competition: Home Credit Default Risk

Summer 2018

- Led team to a top 11% of whole competition with approximately 7100 groups
- Performed EDA and feature importance on datasets
- Developed XGBoost model and tuned parameters in Python to predict results

ACTIVITIES

Pennsylvania State University

ASA DataFest PSU

March 2018-March 2019

- Collaborated with teammates to compete in events where sponsors from organizations such as indeed provide data for student to explore
- Achieved top 5 out of 40 groups in 2018 competition as a member of Team J4 Fun

Available Summer 2021