

Hai Huang

Jersey City, NJ hhuang37@stevens.edu

814-954-9069

<https://www.linkedin.com/in/hai-huang0629/>

EDUCATION

Stevens Institute of Technology, Hoboken, NJ

Master of Science in Data Science

GPA: 3.54

Expected May 2022

Related Courses: Deep Learning, Machine Learning, Statistical Methods, Big Data Technologies, Web Mining

Pennsylvania State University, University Park, PA

Bachelor of Science in Data Sciences

May 2020

Bachelor of Science in Applied Statistics

May 2020

SKILLS

Languages & Skills: Python, SQL, R (R Studio), Deep Learning (Keras), Machine Learning (scikit-learn, LightGBM, XGBoost), Data Visualization (Matplotlib, Plotly, Tableau), Pyspark, DataBricks

WORK EXPERIENCE

Russell Reynolds Associate, New York, NY

January 2022–Now

Data Scientist Intern

- Implement Keras_tuner to hyperparameter tuning LSTM model and improve the performance of the LSTM model to 76% accuracy with 0.58 f1 score
- Performed exploratory data analysis on top 3 most common events, tested significance of new features and applied LDA to perform topic modelling

Russell Reynolds Associate, New York, NY

June 2021 – August 2021

Data Scientist Intern

- Built-up four binary classification models with **TF-IDF**, **Random Forest**, **LDA**, and **LSTM** using **scikit-learn**, **Keras**, and **NLTK** library to predict CEO turnover, which TF-IDF with Random Forest model accomplish **80%** accuracy and presented to director and global head
- Dug deep into the **70K** resume data and conducted exploratory data analysis using **Pandas**, NumPy, NLTK, and Matplotlib to identify trends, patterns, insights and presented the findings to the team
- Applied SQLAlchemy package to connect to SQL server on Jupyter Notebook and used SQL to extract data

Agriculture Bank of China, Guangdong, China

May 2019 – August 2019

Data Scientist Intern

- Discussed and planned for a recommendation system for wealth management products with supervisor
- Employed **LightGBM** model and used **Gridsearch** to tune parameters in Python for an internal machine learning competition and placed No.6 in the competition with 45 other groups (predict how likely customers will purchase products)

PROJECTS

PSU Capstone Project: Estimate the Upper Bound of US Innovation

May 2020

- Collected and cleaned data from multiple datasets and 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 Finish Placement Prediction

December 2019

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

Kaggle Competition: Home Credit Default Risk

Summer 2018

- Achieved top **11%** of whole competition with approximately 7100 groups
- Performed EDA and feature importance on datasets using Python with Pandas and XGBoost packages
- Developed **XGBoost** model and tuned parameters in Python to improve accuracy

ASA DataFest PSU

March 2018

- Achieved finalist award in [ASA DataFest PSU 2018 competition](#) with seventy participating teams
- Worked in a team to use **Tableau** to create various plots to analyze 2.59GB of data provided by **Indeed**