

Hai Huang

Jersey City, NJ | hhuang37@stevens.edu | 814-954-9069 | [LinkedIn](#) | Available Summer 2022

EDUCATION

Stevens Institute of Technology, Hoboken, NJ Expected May 2022

Master of Science in Data Science GPA: 3.54

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

Pennsylvania State University, University Park, PA May 2020

Bachelor of Science in Data Sciences & Bachelor of Science in Applied Statistics

SKILLS

Technical Skills: Python, SQL, R, Pyspark, Machine Learning (sklearn, XGBoost, LightGBM), Deep Learning (Keras), Data Visualization (Matplotlib, Seaborn, Plotly, Tableau)

EXPERIENCE

Russell Reynolds Associate, New York, NY 01/2022 - Present

Data Scientist Intern

- Performed exploratory data analysis, tested significance of new features and applied LDA to perform topic modeling, implemented hyperparameter tuning for LSTM models using Keras
- Took in LDA transformed data and applied Random Forest and Gradient Boosting Classifier models to make classification
- Analyzed client feedback data and build models to classified feedbacks into different categories

Russell Reynolds Associate, Remote 06/2021 – 08/2021

Data Scientist Intern

- Built up classification models with TF-IDF, Random Forest, LDA and LSTM using sklearn and Keras library to predict CEO turnover, which TF-IDF model accomplished 80% accuracy with a 0.54 F1 score
- Performed exploratory data analysis on **70K** Resume data and generated visualizations using Python pandas, matplotlib, and seaborn packages in both Jupyter and Zeppelin notebooks
- Constructed data transformation pipelines from SQL in Python using SQLAlchemy and sklearn to extract CapIQ data from Microsoft SQL Server

ACADEMIC PROJECTS

Netflix Movie Recommendation Fall 2021

- Built ALS recommendation model using Pyspark with Pyspark.ML and mlflow packages on Databricks, achieve 1.24 RMSE score
- Performed data cleaning and data aggregation using Python with Pandas and NumPy packages

PSU Capstone Project: Estimate the Upper Bound of US innovation Spring 2020

- Collected and performed data cleaning from multiple datasets and performed exploratory data analysis on region, income, education level and numbers of patent
- Identified the flaws of given data and convinced sponsors to change and let team to find new datasets

PUBG Finish Placement Prediction Fall 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

COMPETITIONS

Kaggle Competition: Home Credit Default Risk Summer 2018

- Achieved top 11% of whole competition with approximately 7100 groups
- Performed EDA, feature importance, feature selection on datasets using Python with XGBoost package
- Implemented XGBoost model and tuned hyperparameters in Python to improve accuracy

ASA DataFest PSU March 2018

- Formed team J4Fun and achieved finalist award in [ASA DataFest PSU 2018 competition](#)
- Used Tableau to create data visualization and analyze 2.59GB of data provided by **Indeed**