Di Wu

410-600-5808 | dwu49@jh.edu
909 W University Pkwy, MD 21210

Resume Objective: Data Science Master student in Johns Hopkins University with passion and previous work experience in data analysis; Seeking a **Data Science Internship**; Able to commit at least 16 weeks.

SKILLS

- **Programming:** Python, R, JavaScript, Java
- **Big Data:** Hadoop, Hive (HQL)
- **Database:** MySQL, PostgreSQL, MongoDB
- **ML Algorithm:** Decision Tree, Random Forest, Support Vector Machine, K-Means, KNN, Linear Regression, Neural Network, Natural Language Processing, PCA
- **Deep Learning:** PyTorch
- **NLP:** HMM, RNN, N-gram
- **Computer Vision:** CNN
- **Framework:** React, Angular
- **Visualization:** Excel, Tableau, Google Sheet
- **Linux and Shell**
- **Data Crawling Method:** Selenium, requests
- **CDA** (CERTIFIED DATA ANALYST CERTIFICATE)
- **Web:** HTML5, CSS, Node.js

EDUCATION

MSE, Data Science
Johns Hopkins University | Baltimore, MD | Dec 2022

BS, Data Science, Software Engineering
University of California | Irvine, CA | Dec 2019
Overall GPA: 3.6/4.0
Dean's Honor List | Dec 2018 – Jun 2019

LEADERSHIP

Build Web Applications | Dreams for Schools | Tustin, CA | Jan 2019 – June 2019
[Collaborated with four team members to build two grading applications]

- Coordinated team members; Allocated time reasonably; Collected feed back in time
- Set up detailed meeting plan before each meeting; Organized multiple rehearsals before team presentation
- Participated in two coding competitions for 5K+ students
- Launched two applications successfully on host page of company

WORK EXPERIENCE

Data Analyst Intern | Beijing Kuaishou Technology Co. | Beijing, China | May 2021 – Aug 2021
[Responsible for data analysis in social relations direction of Kuaishou application]

- Conducted attribution analysis on key indicators of social relations concern with hundreds of million data by Hive SQL and Statistical Inference; Output weekly attribution reports
- Explored content of group chat information of different groups; Discovered whether user status contains information indicating mood by Latent Dirichlet Allocation
- Made decisions for AB Test results of different scenes of Kuaishou application with different statistical tests to ensure experiments can be implemented
- Achieved 70% user penetration rate on Follow Page; Increased Consumption time by 3%; Decreased zero play rate by 5%

Data Analyst | Er pai Technology Media Co. | Hubei, China | Aug 2020 – Jan 2021
[Responsible for data analysis of company's main product "K Planet" application]

- Wrote automated data crawling and data mining system with Python selenium; Performed data analysis and data mining on backend website and Umeng (data platform website); Transformed user log behavior data generated in app product into relational database structure
- Inspected, added, and deleted product buried points based on application updating; Operated data mining and exploration of buried points to optimize products
- Analyzed user behaviors data via topic modeling; Produced visual data analysis reports regularly
- Won Golden Rice Award from Xiaomi platform in November; Increased active users by 40% in fourth quarter

Data Analyst Intern | Hubei Daily Media Group | Hubei, China | May 2020 – July 2020

- Involved in Big Data project to maintain database, and configure Hadoop environment in Linux
- Obtained data using Elasticsearch and Kibana
- Recorded each processing step for future maintenance

PROJECTS

Obesity Level Analysis Project | JHU | Baltimore, MD | Oct 2021 – Dec 2021

- Discovered relations between obesity level and people's eating habits as well as physical conditions
- Explored original dataset; Checked null value; Checked target value balances
- Performed feature engineering to categorical features and normalized numerical features
- Created different machine learning models including developed super learner; Conducted hyperparameter tuning and model evaluation
- Accomplished 87% test accuracy on tuned super learner model for predicting people's obesity level

Movie Box Office Prediction Project | UC Irvine | Irvine, CA | Oct 2019 – Dec 2019

- Predicted movie box offices by using 26 independent variables including YouTube trailer comments and Twitter feedback comments
- Accomplished data crawling from internet including YouTube comments, like and subscription, Twitter followers of director and comments after one week of film's release
- Stored crawling data into PostgreSQL database; Developed different Machine Learning models
- Reached 85% accuracy for best model