Di Wu

410-600-5808 | dwu49@jh.edu | LinkedIn
1269 Commonwealth Ave, MA, 02134

Resume Objective: Data Science Master student at Johns Hopkins University with passion and previous work experience in data science and machine learning; Seeking a **Data Scientist Full Time Role.**

EDUCATION

MSE, Data Science
Johns Hopkins University | Baltimore, MD |
Dec 2022 | Overall GPA: 3.5/4.0

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

SKILLS

- **Programming:** Python, R, JavaScript, Java
- **ML Algorithm:** Decision Tree, Random Forest, Support Vector Machine, K-Means, KNN, Linear Regression, Logistic Regression, PCA, Ensemble
- **NLP:** N-gram, RNN, GRU, LSTM, Transformer, BERT
- **Computer Vision:** CNN, ViT, GANs, YoLo V3, Swin-Transformer
- **Big Data:** Hadoop, Hive
- **Database:** MySQL, PostgreSQL, MongoDB
- **Framework:** Pytorch(DL), React, Angular
- **Visualization:** Excel, Tableau, Google Sheet
- **Linux | Shell | Slurm**
- **Data Crawling Method:** Selenium, requests
- **CDA** (CERTIFIED DATA ANALYST CERTIFICATE)
- **Web:** HTML5, CSS, Node.js

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 Scientist Intern | Boston Children's Hospital, Harvard Medical School | Boston, MA | July 2022 – Dec 2022
[Responsible for HIE and Nutrition outcome Prediction, MRI Image Segmentation]
- Collaborated with cross-functional teams (MD students and professors) to determined general clinical data workflow pipeline
- Built new FIBE (Forward Inclusion Backward Elimination) method for feature selection; Supported grouped features to enhance interpretability for one-hot encoding features
- Replaced 3D U-net approaches with 3D U-net Transformer for MRI segmentation task; Combined MRI (image) data with clinical (tabular) data, which increased outcome prediction accuracy to 82%
- Performed 20+ experiments dealing with unbalanced data, metrics selection, missing values, and hyperparameters-tunning; Increased 15% accuracy than baseline models

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
- Explored content of group chat information; Discovered whether user status contains information indicating their sentiments by Latent Dirichlet Allocation
- Made decisions based on A/B 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 configured Hadoop environment in Linux
- Obtained data using Elasticsearch and Kibana
- Recorded each processing step for future maintenance

PROJECTS

Image Captioning (Capstone Project) | JHU | Baltimore, MD | Sep 2022 – Dec 2022
- Utilized Seq2Seq model structure to develop pretrained CNN + GRU models as baseline model for image captioning tasks on Flickr8k dataset with 47.19 BLEU score
- Added attention mechanism on decoder to reduce encoder's burden which Increased BLEU score to 55.65
- Established new two-stage transformer model that took advantage of Vision Transformer, a simple and small CNN network, Image augmentation, and Beam Search which reaches to a 61.55 BLEU score. The new model has higher BLEU score than pretrained model (ViT + GPT-2) from Hugging Face (59.88 BLEU score)