Mingxuan (Curly) Wu

Email: curlywu99225@gmail.com | Mobile: (+1)858-295-9303 | Website: https://www.linkedin.com/in/curly-wu

EDUCATION

New York University, Center for Data Science (CDS) Sep 2022 – Dec 2023
Candidate for the **M.S. in Data Science**

University of California, San Diego (UCSD), CA Sep 2018 – Mar 2022
B.S. in Data Science and **B.S. in Management Science**
Overall **GPA of 3.91 / 4.0**
Cum Laude Distinction, UCSD Provost Honor for 11 Quarters, Member of Phi Beta Kappa Sigma Chapter of California

WORK EXPERIENCES

Data Scientist Intern | Franklin Templeton Investment Jun 2022 – Aug 2022

- Extracted-transformed-loaded FRED data, **leveraged time series analysis** to forecast the Inflation rate, handled **post-pandemic data abnormality** and achieved a **45.6% improvement in MAE**:
  - Examined and transformed 50+ macroeconomic variables based on **ADF tests** scores, filtered out 30+ variables that linearly aligned with the target using regression P-values, and dropped highly correlated variables using **VIF tests**
  - Tracked data abnormality during the pandemic and aligned the data by capping outliers and feature selections
  - **Time Series Spited data** and **tuned hyperparameters** of models including **Lasso, Ridge, Random Forest, and Light GBM**, achieving a 45.6% improvement in MAE over the **empirical Baseline**
  - Visualizing modelling results and creating **workflow charts** for a data science presentation to business users

Data Analyst Intern | JDD Tech Company Jul 2021 – Sep 2021

- Assisted the collection and analysis of customer behaviors data**:**
  - Processed **high volumes of customer data** to capture user persona at JDD in support of digital marketing strategies
  - **Transformed Flume HDFS Interceptor to Flink HDFS Interceptor** to improve data throughput by 200%
  - Leveraged 10+ software, including **GitHub**, **Hive**, IDEA, **MySQL**, **Spark** Core, Flume, Flink, Navicat, Kafka, and Scala language to accomplish the tasks; performed batch job scheduling using Azkaban Hadoop

Data Analyst Intern | Alibaba Group Jun 2020 – Aug 2020

- Completed a data-driven comparative study on the topic of live commerce vs. traditional e-commerce in China:
  - Performed **data mining** and **analyzed 6+ forms of KPIs** such as page view (**PV**), unique visitor (**UV**), daily active user (**DAU**), monthly active users (**MAU**), peak concurrent users (**PCU**), average revenue per user (**ARPU**), etc.
  - Queried and organized large-scale customer data with the aid of SQLite relational database
  - Conducted customer review analysis and **sentiment analysis** via cross-validated **random forest, gradient boosting, logistic regression** with texts **TF-IDF encoded**, and Naïve Bayes models

PROJECTS

Team Leader | Blockchain and Smart Contract Application: Gym Coin
Data Science Senior Capstone Design, UCSD Sep 2021 – May 2022

- Developed a **Blockchain**, **Smart Contract**, and **NFT** application for commercialization of daily exercises:
  - Researched on cryptocurrency tokens (ERC20 and ERC721) and their decentralized applications, developed an innovative solution that decentralizes the exercise-reward systems, and synchronize the ideas into a whitepaper
  - **Developed** and **tested** the smart contract on **Remix** using **Solidity**, implemented the contract on Ethereum ropsten test network, and deployed and shipped on Scaffold-ETH for front-end User Interface: http://dsc180a03_gymcoin.surge.sh/

Team Leader | Business Analytics for GoShare Sep 2020 – Jan 2021

- Established a data-driven workflow to support the last mile logistics platform at **GoShare**:
  - **Enabled automated data mining** from the Metabase business intelligence tool at GoShare, focusing on evaluating project completion & cancellation rates and service ratings while considering geospatial and partnership data
  - Implemented **descriptive data analytics** and **visualization** on the collected data to drive performance evaluation for GoShare employees
  - Performed **diagnostic and predictive data analytics** to forecast the acceptance rates, incorporating the rate of return and credit scores of customers into the models to optimize resource allocation and pricing

Team Leader | Cloud/Cluster Computing Course Project, UCSD Sep 2020 – Dec 2020

- Built a **million-scale data-centric pipeline on AWS** to predict product ratings and **optimize marketing strategies**:
  - Processed more than **45 million** data points by cleansing and **flattening** the semi-structured data, **imputing** missing data, and conducting **PCA** to reduce dimensionality redundancy, and applying **one-hot encoding** to handle sentiments
  - Trained and validated supervised models such as **logistic regression**, **random forest**, and **decision tree** to drive predictions
  - Conducted hyperparameter tuning and cross-validation to **boost accuracy to 81.5%**

SKILLS

Programming Languages: Python (**NumPy**, **Pandas**, **Scikit-Learn**, **PySpark**, **Matplotlib**, Altair), **SQL, R** (ggplot, dplyr), Java, MATLAB, Stata, Scala, Dask.
Software: Jupyter Notebook, Git, IntelliJ, Microsoft Office, Visual Studio, PyCharm, Eclipse.