Xinyang(Oliver) Zhou

xinyangzhou2024@u.northwestern.edu | 734-510-0762 | linkedin.com/in/xinyang-zhou-45b95a218

EDUCATION

Northwestern University **Evanston, IL**
Master of Science in Machine Learning and Data Science; Cumulative GPA: 3.9/4.0 Dec 2024
- Relevant Coursework: Big Data Analytics, Data Warehousing, Cloud Engineering, Text Analytics (NLP)

University of Michigan **Ann Arbor, MI**
Bachelor of Science in Statistics, Minor in Mathematics and Business; Cumulative GPA: 3.84/4.0 May 2023
- Scholarship: Veeam Software Endowed Scholarship in Data Analytics, 2021-2023 Academic Years
- Awards/Honors: James B. Angell Scholar; University Honor for All Semesters Enrolled

TECHNICAL SKILLS

Programming: Python (ex. scikit-learn, NumPy, pandas, TensorFlow, Matplotlib, PyTorch, PySpark, etc.), R, SQL, SAS, Java, C++
Relevant Platforms: GitHub, AWS, Google Cloud Platform, VS Code, Jupyter Notebook, Snowflake, DataGrip, SQLite
Software: Tableau, Power BI, Microsoft Office, Hadoop, Spark, Hive, Hbase, Docker
Campus Involvement and Leadership: Ross School of Business: Finance Course Instructor; Linear Algebra Course: TA/Grader; University of Michigan Chinese Soccer Team: Captain; ImmersiveLIVE Drama Club: Starring/Associate Director

WORKING EXPERIENCE

CDK Global **Greater Chicago Area, IL**
Software Engineer Intern (Data Science focused) Jun 2024 – Aug 2024
- Built predictive buying models and streamlined MLOps (deployment, automation, cloud integration, etc.)
- Created suggested vehicle models that provide marketing tools for dealerships to solicit sales and capture/retain customers

Little City Foundation **Greater Chicago Area, IL**
Data Science Consultant Sep 2023 – Jun 2024
- Performed **feature engineering** to create responsive prospecting models (**classification**, **regression**, and **clustering**) that predict future donor categories/levels based on 25 years of data to facilitate revenue growth
- Identified outliers/influential points and enhanced models to find the optimal one by using different **variable selection methods** (Backward/Forward Selection, Regularization), enhanced performance by 30%
- Constructed machine learning models in **Python** to investigate the previous donor behaviors and maintained 90%+ high-level donors; expanded mid-level donor portfolio by 56%

Nationwide Mutual Insurance Company **Columbus, OH**
Business Analyst Intern May 2023 – Aug 2023
- Designed and implemented the **A/B testing** to investigate the root cause of the difference in service accuracy between markets and request types, boosted the productivity within the organization by 15%
- Incorporated Nationwide and Allstate datasets and created predictions of the combined loss ratio for 50 states of the National Retail Programs by using **SQL** on **Snowflake**; collaborated with the underwriting team to make decisions
- Visualized the percentage difference between technical premium and premium in force in 41 states in the U.S. using **Tableau** to support pricing strategy in 30 states; aided 70% revenue growth nationwide and minimized risks
- Analyzed 1M rows of cognizant quality data and explored potential influences of processing time and number of touches on error rate, guiding senior VPs to adjust relevant policies

Michigan Institute for Data Science **Ann Arbor, MI**
Data Analyst Assistant Apr 2022 – Apr 2023
- Generated **data visualizations** using **Excel** and **R** (dplyr, ggplot2, etc.), enabling data visibility in the esports and entertainment industries to facilitate the data mining process in the research
- Conducted **data mining** on the dataset (100k rows, 25 columns) to pinpoint **causal inference** for the team's performance, local crime rate, and team franchise value, visualizing unsupervised learning results using Yellowbrick

RESEARCH PROJECT

Predictive Model Analysis for Profit Optimization Strategy in Telemarketing Centers of Banks (R)
- Developed models using **logistic regression**, **neural networks**, **KNN**, **SVM**, **decision tree**, **random forest**, and **Gradient Boosting Tree** to predict the telemarketing call result, tuned each model by **cross-validation(CV)**
- Linked to the business problem and implemented **loss matrix/function** to the previous models and produced a final deliverable containing qualitative contents analysis (**Precision**, **Recall**, **F1 Score**) along with the proof from modeling in a generally understandable tone to indicate the best **stacked model**

Future Stock Price Prediction of the Goldman Sachs Group Using Deep Learning Algorithms (Python)
- Designed 8 models in different structures to predict the future stock value of Goldman Sachs including **LSTM**, **SimpleRNN**, and **Conv1D Layer**, fine-tuned the model by adjusting hyperparameters (learning rate, etc.)
- Interpreted the performance and updated the model while maintaining a balance with computational complexity
- Achieved an accuracy >98.5% for the best model based on a dataset containing daily market information from May 1999 to March 2022 (7 columns and 5,000+ rows)

Spotify Song Recommendations (Python)
- Implemented dimensionality reduction models using **t-SNE** and **PCA**, optimized by elbow plot
- Selected the best clustering model from **K-Prototype**, **K-Means**, and **DBSCAN** based on the **Silhouette** score
- Finalized a hybrid **recommendation system** that feeds from item-based and content-based recommendations, optimizing the best weights by minimizing the RMSE; designed a user interface for song recommendations