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: 3.933/4.0 Dec 2024
- Relevant Coursework: Cloud Engineering, Text Analytics (NLP), Deep Learning, Optimization, Reinforcement Learning

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/Honors: Veeam Software Endowed Scholarship in Data Analytics; James B. Angell Scholar; University Honors

TECHNICAL SKILLS

Programming: Python (ex. scikit-learn, NumPy, pandas, TensorFlow, PyTorch, PySpark, etc.), R, SQL, SAS, Java, C++
Relevant Platforms: GitHub, AWS, Google Cloud Platform (GCP), Snowflake, PostgreSQL, VS Code, Jupyter Notebook, PyCharm
Software: Tableau, Power BI, Microsoft Office, Hadoop, Spark, Hive, HBase, Docker, Airflow, Confluence, Jira, dbt
Campus Involvement and Leadership: Ross School of Business: Finance Course (FIN 302) Instructor; Convident (student start-up, language learning app): Staff Machine Learning Engineer; Northwestern Medicine: Research Assistant for LLM and NLP; Harvard Business School: Research Assistant for LLM and A/B testing

WORK EXPERIENCE

Amazon **Greater Chicago Area, IL**
Data Science Consultant Sep 2024 – Dec 2024
- Developed a hybrid fault detection system combining **NLP** for unstructured text and **time series analysis** for structured monitoring data, reducing customer impacts and improving network reliability
- Used **Word2Vec** embeddings and **cosine similarity** to measure document similarity, and applied **Dynamic Time Warping (DTW)** to identify network event patterns and enhance accuracy in detecting recurring network issues

CDK Global **Greater Chicago Area, IL**
Software Engineer Intern Jun 2024 – Aug 2024
- Generated propensity scores using **ML** models and followed **MLOps** principles, enhancing **AI marketing** efficiency
- Utilized **SQL** on **Snowflake** to detect and fix errors in the mapping algorithm, increased customer retention rates by 10%
- Autonomized the **ELT/ETL** process of dealership data through **Airflow**, minimized data quality error rate by 15%

Little City Foundation **Greater Chicago Area, IL**
Data Science Consultant Sep 2023 – Jun 2024
- Performed **feature engineering** to create predictive models (**XGBoost**, **Lasso**, **GLM,** etc.) for future donor categories/levels; suggested **customized events** to donors, receiving a 25% increase in satisfaction
- 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 **ML pipeline** and utilized **Docker** to ensure reproducibility and consistency

Nationwide Mutual Insurance Company **Columbus, OH**
Business Insights Analyst Intern May 2023 – Aug 2023
- Incorporated Nationwide and Allstate datasets, analyzed the financial status and prospective returns, and created predictions of the combined loss ratio using **SQL** on **Snowflake**; collaborated with the **underwriting** team to make decisions
- Designed and implemented **A/B testing** to investigate the root cause of the difference in service accuracy between markets and request types by **statistical hypothesis tests** in **Python**, boosted productivity by 15%
- 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 3.7% revenue growth and minimized commercial line **financial 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 **pricing strategy** using **pivot tables** and **charts**

Michigan Institute for Data Science **Ann Arbor, MI**
Data Analyst Assistant Apr 2022 – Apr 2023
- Generated **data visualizations** and **augmentation** using **R** (dplyr, ggplot2, labeling, etc.), enabling data visibility in the esports and entertainment industries to facilitate the data mining process
- 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 results using **Matplotlib**, **Seaborn**, and **Plotly**

RESEARCH PROJECT *(GitHub Link)*

Predictive Model Analysis for Profit Optimization Strategy in Telemarketing Centers of Banks
- Developed models using **logistic regression**, **neural networks**, **KNN**, **SVM**, **decision tree**, **random forest**, and **Gradient Boosting Tree** to predict the telemarketing call result, fine-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; awarded by the industry sponsor (top 3)

Spotify Song Recommendations
- Implemented dimensionality reduction models using **PCA**, optimized by elbow plot; visualized data by using **t-SNE**
- 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