

Xinyang(Oliver) Zhou

xinyangzhou2024@u.northwestern.edu | 734-510-0762 | [linkedin.com/in/xinyang-zhou-45b95a218](https://www.linkedin.com/in/xinyang-zhou-45b95a218)

EDUCATION

Northwestern University

Master of Science in Machine Learning and Data Science; Cumulative GPA: **3.93/4.0**

- Relevant Coursework: Big Data Analytics, Data Warehousing, Cloud Engineering, Text Analytics (NLP)

University of Michigan

Bachelor of Science in Statistics, Minor in Mathematics and Business; Cumulative GPA: **3.84/4.0**

- Scholarship: Veeam Software Endowed Scholarship in Data Analytics, 2021-2023 Academic Years
- Awards/Honors: James B. Angell Scholar; University Honor for All Semesters Enrolled

Evanston, IL

Dec 2024

Ann Arbor, MI

May 2023

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), 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 (FIN 302) Instructor; Linear Algebra Course (MATH 214): TA/Grader; University of Michigan Chinese Soccer Team: Captain; ImmersiveLIVE Drama Club: Starring

WORKING EXPERIENCE

CDK Global

Software Engineer Intern (Data Science focused)

Greater Chicago Area, IL

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

Data Science Consultant

Greater Chicago Area, IL

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

Business Analyst Intern

Columbus, OH

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

Data Analyst Assistant

Ann Arbor, MI

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 results using **Matplotlib**, **Seaborn**, and **Plotly**

RESEARCH PROJECT

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, 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

Animal Image Classification and Emotion Detection

- Designed models with **ImageDataGenerator**, **early stopping**, and **learning rate scheduler** to boost performance
- Incorporated hyperparameter tuning to find the best parameters and utilized **Tensorboard** to monitor and visualize
- Applied **transfer learning** including InceptionV3, **YOLO**, MobileNetV2, and VGG with customized layers to build high-performance models (bounding box accuracy, loss, etc.)

Spotify Song Recommendations

- 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