Xinyang(Oliver) Zhou

xinyangzhou2024@u.northwestern.edu | 734-510-0762 | linkedin.com/in/oliverzhou-data-science-analytics

EDUCATION

Northwestern University **Evanston, IL**
M.S. in Machine Learning and Data Science; Cumulative GPA: 3.94/4.0 Sep 2023 - Dec 2024
- Relevant Coursework: Generating Business Values by Analytics, NLP, Deep Learning, GenAI

University of Michigan **Ann Arbor, MI**
B.S. in Statistics, Minor in Mathematics and Business Aug 2019 - May 2023
- Scholarship/Honors: Veeam Software Endowed Scholarship in Data Analytics; James B. Angell Scholar

TECHNICAL SKILLS *(GitHub Link)*

Programming: Python (Pandas, NumPy, Matplotlib, scikit-learn, TensorFlow, PyTorch, PySpark, etc.), SQL, R, C++
Data Management and Deployment: AWS, GCP, Snowflake, Docker, Airflow, dbt, Hadoop, Spark, Hive
Analytics and Visualization: Tableau, Power BI, ggplot2, Matplotlib, Seaborn, Looker
Platforms and Tools: GitHub, Confluence, Databricks, PostgreSQL, Jira, Notion

WORK EXPERIENCE

Freddie Mac **McLean, VA**
Quantitative Analyst Apr 2025 – Present
- Built a house condition prediction model using **LightGBM**, reaching an AUC>90%, facilitating overall valuation
- Performed comprehensive **model validation**, including out-of-time testing, sensitivity analysis (shock scenarios), and **k-fold cross-validation** to ensure performance stability and robustness
- Designed and implemented **custom model explainability metrics** (SHAP values), and presented **audit-ready reports** to both technical reviewers and business stakeholders for regulatory compliance

Amazon **Chicago, IL**
Data Science Intern - Capstone Sep 2024 – Dec 2024
- Applied **NLP** techniques to analyze system fault logs and engineered sequence embeddings using **LSTM**, improving failure pairing speed and efficiency in AWS infrastructure diagnostics
- Designed a reproducible reasoning pipeline with full version control on **GitHub**, enhancing model traceability and reducing engineering **turnaround time** by 15% through structured experimentation

CDK Global **Chicago, IL**
Software Engineer Intern Jun 2024 – Aug 2024
- Productionized an **unsupervised** algorithm to generate propensity score, improving targeting logic accuracy
- Utilized **SQL** on **Snowflake** to fix errors in the mapping algorithm, increased customer retention rates by 10%
- Collaborated with product managers to integrate user feedback and define **business thresholds** for monitoring

Little City Foundation **Chicago, IL**
Data Science Intern - Practicum Sep 2023 – Jun 2024
- Conducted **observational studies** to analyze donor behavior and identify satisfaction drivers, applying statistical methodologies and providing data consulting to ensure stakeholder understanding and adoption
- Engineered features and built **predictive models** (XGBoost, Logistic) to segment donors and forecast donation levels; employed **difference-in-differences** analysis to estimate the impact of targeted interventions
- Constructed **ML pipeline** and utilized **Docker** for deployment, ensuring reproducibility across analyses

Nationwide Mutual Insurance Company **Columbus, OH**
Business Insights Analyst Intern May 2023 – Aug 2023
- Designed **A/B tests** to identify key factors influencing service accuracy across regions, leading to a 15% productivity increase; conducted initial **power analysis** to determine test duration and sample size
- Created **interactive dashboards** in **Tableau** and **Power BI** across 41 U.S. states, informing pricing strategies and contributing to a 3.7% increase in revenue while mitigating financial risk
- Implemented complicated SQL queries (**CTE**, **window function**, etc.) to support downstream analytics

RESEARCH EXPERIENCE

Harvard University, Harvard Business School **Remote**
Research Assistant / Co-Author
- Designed and implemented a modular ML pipeline for tabular data processing, SMOTE for class imbalance, covariate selection, and prediction generation using **LLM APIs** with **prompt engineering** strategies (few-shot)
- Enhanced large language model (LLM) performance through **fine-tuning** and prompt optimization; visualized model outputs and presented **LLM-driven marketing use cases** at an academic conference
- Defined and tracked **custom business metrics** to evaluate LLM reliability compared to traditional ML models; applied **causal inference methods** (e.g., average treatment effect, statistical significance testing) to uncover robust insights