Lingyi Zhao

lingyiz2@uchicago.edu | Lingyi-Zhao-01 | 217-305-0295 | Chicago, IL

EDUCATION

University of Chicago — **Chicago, IL**
Master of Science in Analytics — *Aug 2020 - Mar 2022*
- Big Data Platforms, Machine Learning, Data Mining, DS for Consulting, Data Engineering Platforms, Time Series

University of Illinois at Urbana-Champaign — **Champaign, IL**
Bachelor of Science in Econometrics and Quantitatives (*Minor in Business and Statistics*) — *Aug 2016 - May 2020*
- Vice President of Publicity and Relations at Chinese Students And Scholars Association (CSSA)

SKILLS

Programming: R, Python, SQL, Git, Spark, Hadoop, OpenRefine, Neo4j, MongoDB, Tableau
Libraries: Pandas, NumPy, ggplot, Matplotlib, Scikit-learn, SciPy, OpenCV, TensorFlow, PyTorch, Keras, NLTK, Flask

WORK EXPERIENCE

W.R.Berkley - Small Business Solutions — **Wilmington, DE**
Data Science Intern — *Jun 2021 - Sep 2021*
- Extracted insurance transactions from **SQL Server**, transformed JSON data-payload and flagged anomalous transactions in **Python**
- Conducted feature-level analysis on model score variation, designed a 3-layer alarm strategy to avoid false positive-alarms
- Implemented a real-time anomaly detection **Flask** service and testing through **Postman**, sending emails to agents and stakeholders
- Deployed the service on **Kubernetes** and applied the service company-wide, successfully catching abnormal cases with 95% accuracy
- Documented project's design in Confluence through **JIRA** and controlled code version through **Git**, improving accessibility in the team

University of Chicago — **Chicago, IL**
Data Analysis Intern — *Mar 2021 - Jun 2021*
- Created monthly HR metric reports by gathering, validating, and analyzing data from core systems, visualized results using **Tableau**
- Performed statistical analysis in **Excel** to assist HRBP in determining human capital efficiency and salary range
- Applied data-orientated improvements for the HR daily working, enabled **Python** code to help reduce 5 hrs task into 30 mins

Deloitte Consulting — **Shanghai, China**
Data Analytics and Information Management Intern — *May 2018 - Aug 2018*
- Performed on a data architecture and visualization project for a bank loan department to optimize database management system
- Designed a new data naming system to solve the problem of inconsistency by standardizing 2k+ metrics and 3k+ dimensions
- Constructed a relational database using **SQL** to improve the capacity and efficiency of the new business intelligence system
- Visualized data by creating a reporting dashboard on **Tableau** to speed up the decision making process

ANALYTICAL PROJECTS

B2W Marketplace Seller Future Value Prediction (Capstone Project) — *Mar 2021 - Present*
- Conduct meetings with top management and cross-functional teams, devising growth strategies and ensured our result met GMV targets
- Extract sales data using **GCP** and **Google BigQuery**, analyzing metrics and segmenting sellers into groups based on **RFM** score
- Monitor prediction results from time series(**ARIMA**) and machine learning models(**Regression/ Random Forest/ Gradient Boost**), adjusting feature weights overtime to get a comprehensive model
- Deliver seller future value prediction results on an established Tableau dashboard, enabling non-technician to easily reachable

Banking Customer Churn Prediction Analysis — *Mar 2021*
- Developed algorithms for telecommunications service vendors to predict customer churn probability based on labeled data via Python
- Preprocessed data by data cleaning, categorical feature transformation and standardization
- Trained supervised ML models including **Logistic Regression**, **Random Forest** and **KNN**, and applied regularization with optimal parameters to overcome overfitting
- Evaluated model performance via cross-validation and analyzed feature importance to identify top factors that influenced the results

NLP and Topic Modeling on Amazon User Review Dataset — *Feb 2021*
- Clustered customer reviews into groups, preprocessed text by tokenization, lemmatization, and extracted features by TFIDF
- Trained unsupervised learning models using **K-means** and **LDA**, identified latent topics and keywords for clusters
- Used topic results to perform supervised learning by adding features, predicted Amazon's customer rating and generated recommendations

Lego Database Analysis — *Oct 2020 – Dec 2020*
- Constructed a relational database using extracted and normalized 300K+ Lego's historical inventory and sales data (**OpenRefine**)
- Designed a Tableau dashboard and connected to the SQL database to monitor major KPIs from 8 dynamic graphs
- Made marketing recommendations by forecasting future price,segmenting customer groups and analyzing historical patterns vs current trends at set/color/theme level