

Xinyi (Eve) Sheng

984-377-0506 | evexyin@gmail.com | [linkedin.com/in/xinyi-sheng/](https://www.linkedin.com/in/xinyi-sheng/) | Durham, North Carolina

Education

Duke University

09/2022-05/2024 (Expected)

M.S in Statistical Science, GPA: 3.85/4.0

Relevant Coursework: Probabilistic Machine Learning (Python), Bayesian Statistical Modeling (R), Deep Neural Nets (Python), Hierarchical Models (R), SQL: Data Query & Analysis

Beijing Normal University

09/2018-06/2022

B.S in Mathematics and Applied Mathematics, GPA: 3.70/4.0

Awards: 1st Scholarship for Outstanding Academic Performance, BNU (awarded to top 10%)

Relevant Coursework: Partial Differential Equations, Mathematical Modeling (Matlab), Time Series Analysis

Skills

Languages and Tools: Python (pandas, numpy, scikit-learn, pytorch), SQL, R, Git (version control), SAS

Experience

Travelers

Hartford, CT, USA

Data Science Intern

06/2023-08/2023

- Performed data extraction and cleaning from corporate databases with SQL and utilized chi-square test and XGBoost for feature engineering process, reducing features by 93%.
- Designed and integrated a k-means clustering approach into a generalized linear model, resulting in a 2% increase in predictive accuracy of general liability insurance pricing.
- Developed a data visualization map using Python geospatial packages, providing stakeholders with Zipcode-level territorial insights.

Kuaishou Technology

Beijing, China

Data Science Intern

10/2021-02/2022

- Optimized SQL queries for extracting video's hourly, reducing data load by 70% and enhancing overall workflow.
- Developed and implemented a multi-layers DNN to predict trending videos using Python, achieving 90% accuracy and reducing manual work (60% accuracy)
- Deployed the ML model in the recommendation and video ranking platform after AB testing, enabling the Operation team to perform video promos of potential trending videos.

Maimai.cn

Beijing, China

Strategic Analysis Intern

07/2021-10/2021

- Collaborated in a team of 6 to conduct EDA on large datasets and design data visualization dashboards using tableau to update top-ranked companies in different industries.
- Collected users' behavior and profiles on maimai app from inner database (SQL Server) and calculated 20+ user metrics which serve as a source of truth for internal analytics across departments.
- Analyzed user behavior and membership data to segment users and developed a new membership pricing system, improving the membership subscription rate by 5%.

Projects

Child Maltreatment Investigation Project

Fall 2023-Now

Capstone Project, Duke University

- Contributed to a research project with the Child Maltreatment Research Team, identifying the social drivers of child maltreatment from data visualizations and descriptive analysis.
- Built a data preprocessing pipelines using Python, leading long electronic health records to structured data and ensuring reliable and consistent analysis.
- Implemented multilevel linear mixed effect model accounting for both individual and community level factors to predict child maltreatment, resulting in a 20% increase in AUC compared to a generalized linear model.

Morphine Price Analysis Project

Fall 2023

Hierarchical Model, Duke University

- Led the analysis of morphine price using R and imputed 10% missing data from resident surveys utilizing Multivariate Imputation by Chained Equation (MICE).
- Constructed multiple linear mixed effect models, addressing the heterogeneity in location by adding the random intercepts.
- Generated a R markdown to present findings and methodology in a detailed report.