

Xingye Zhang (Austin)

Email: azserein1016@outlook.com
Github: www.github.com/az-serein
Linkedin: www.linkedin.com/in/az-serein

Education

University of Connecticut Bachelor of Arts in Statistics and Economics Sep 2020 – Dec 2024

- GPA: 3.88/4.00
- 2024 New England Scholar
- Four Dean's List (Fall 2022 - Spring 2024)

Research Experience

Department of Statistics, University of Connecticut, Storrs, CT Mar 2024 - Present
Research Assistant *Supervised by Professor Mary Lai O. Salvana*

- Conducted preliminary research for city-scale and building-scale disaster prediction by estimating building heights, reprocessing the UT-GLOBUS dataset, developing interactive maps, and performing exploratory spatial data analysis.
- Preprocessed satellite data with QGIS and SNAP, combined with other datasets including OpenStreetMap building footprints, and applied Random Forest Regression to predict urban building heights for disaster modeling.
- Reprocessed data of 413 major U.S. cities (gpkg files) from UT-GLOBUS datasets using HPC, calculating coordinates for each building and exporting attributes of each city into individual CSV files.
- Developed an interactive map using data from New Haven to visualize building footprints and associated attributes. Integrated exploratory spatial data analysis across 4 cities to examine spatial autocorrelation and enhance visualization insights.
- Created a detailed tutorial to ensure reproducibility of data preprocessing, integration, and analysis workflows.

Internship Experience

GBCS - SkyIT Services, Alberta, Canada May 2024 - Aug 2024
Express Backend Developer

- Developed backend services for the Voop application using Express.js and Django REST, contributing to scalable architecture and API integration.
- Collaborated with teammates in migrating database from Firebase to MySQL, optimizing database schema, and ensuring data accuracy during the transition.
- Built and tested APIs for Voop, utilizing Postman to debug and validate functionality.

Qian Shi Du, Beijing, China Jan 2024 - May 2024
Data Analyst and Market Research Associate

Represented Qian Shi Du at 'Vision Expo' in New York City from March 14 - 17.

- Performed time series analysis to compare current and historical trends and developed a GPT-based model to forecast customer demand patterns, with SHAP values used to interpret feature importance.
- Employed K-means to segment customer preferences, guiding targeted product development and marketing.
- Integrated insights from customer segmentation, purchasing behavior, and product tracking to forecast sales trends, supporting the company's strategic decision-making in international trade.

Practical Projects

Bytedance - Data Optimization and Algorithm Enhancement

Remote Collaboration

- Preprocessed large-scale text data, using techniques such as tokenization and stemming to improve data quality and enhance search engine performance.
- Applied a SAITS model for data imputation and fine-tuned a LLaMA2 model by LoRA to predict product prices based on descriptions, addressing missing data challenges.
- Optimized product price prediction algorithms by refining code efficiency and improving data pipelines, resulting in increased accuracy and operational performance.

Exploratory Data Analysis in NYC Open Data—Rodent Inspection

Introduction to Data Science

- Cleaned and analyzed NYC rodent inspection data using methods like binomial regression models.
- Designed an interactive map with GeoPandas to display rodent sighting locations with all related information and displayed the result to the NYC Open Data team to support rodent inspection.

Prediction of NBA Players' Salary based on On-Court Contributions *Introduction to Statistical Learning*

- Cleaned and conducted exploratory data analysis on on-court statistics to examine their correlation with player salaries.
- Utilized 4 machine learning models and applied methods such as hyperparameter tuning, Bayesian optimization, and model ensembling to improve prediction accuracy.

Technical Skills

- **Machine Learning Models, Techniques, and Libraries:**
 - **Supervised Learning:** Random Forest, XGBoost, Lasso, Ridge.
 - **Unsupervised Learning:** Gaussian Mixture Models, K-Means.
 - **Time Series and Natural Language Processing:** ARIMA, LSTM, GPT-2 (Hugging Face), LLaMA2 (fine-tuned using LoRA), SAITS (used for imputation)
- **Programming & scripting:**
 - **Programming Language:** Python, R, SQL (MySQL, SQLite, PostgreSQL), HTML, JavaScript.
 - **Operation System and HPC:** Bash, Linux (Ubuntu, Red Hat), Slurm.
 - **Data Analysis Tools:** Tableau, QGIS, SNAP, SAS, Minitab, Microsoft Excel.

Selected Courses

Statistics and Data Science in Grad-Level

- Applied Spatio-Temporal Statistics, Introduction to Data Science, Introduction to Statistical Learning, Open Source Programming (Python) , Seminar in Statistics

Mathematics and Operation Research

- Calculus 1-3, Linear Algebra, Elem Differential Equations, Operation Research