

# Xinhui Luo

xhluo06@outlook.com | (530)-304-5485 | [LinkedIn](#) | Portfolio: <https://git.xhluo.gd.edu.kg/>  
Enrolled in Tufts Graduate Co-op Program

## SUMMARY

---

I am a data science graduate student with a solid foundation in statistical modeling, machine learning, and data visualization, supported by hands-on experience in Python, R, and SQL. My academic and project work reflects a strong ability to extract insights from complex data—such as building an interactive dashboard for student placement analysis, implementing KNN for text readability, and modeling health outcomes using decision trees and logistic regression. I approach problems with both analytical rigor and practical implementation skills, aiming to deliver interpretable, data-driven solutions that support real-world decision-making.

## EDUCATION

---

### Tufts University

Expected Graduation: 2026

*Master of Science in Data Science*

GPA: 3.5

Relevant Courses: Probability Theory, Statistics, Big Data, Machine Learning, Visual Analytics, Human-Computer Interaction

### University of California, Davis

Graduated: Dec. 2023

*Bachelor of Science in Statistics (Data Science Track)*

GPA: 3.6

Relevant Courses: Linear Regression, ANOVA, Applied Time Series Analysis, Probability Theory, Data Science, Applied Linear Algebra, Calculus, Multivariate Data Analysis, Data and Society, Statistical Data Technologies, Nonparametric Statistics, Statistical Learning, Advanced Statistical Computing, Categorical Data Analysis

## EXPERIENCE

---

### Digital Jinhua Technology Operation Co., Ltd

June 2024 - Aug. 2024

*Data Development Engineer Intern*

Jinhua, China

- Developed multiple dimension tables for data modeling on the DataWorks platform using MySQL.
- Extracted and processed data for Jinhua's Digital Worker Union using Python.
- Designed and developed a data aggregation and exchange platform website using SmarDaten.

### Industrial And Commercial Bank of China, Jinhua Branch

May 2024 - June 2024

*Planning and Financial Department Intern*

Jinhua, China

- Summarized sub-branch financial data using Excel, utilizing pivot tables and VLOOKUP to organize large datasets.
- Performed linear regression analysis using Python to analyze the bank's operating conditions and provide recommendations.

### Zhejiang Yingyang Asset Management Co. LTD

Jul. 2021 - Oct. 2021

*Quantitative Researcher Intern*

Hangzhou, China

- Developed deep neural network models in Python using TensorFlow to predict stock market trends.
- Achieved a 20% profit in one month using a 3-day predictive model based on historical stock data.
- Optimized existing trading models through data preprocessing, feature engineering, and hyperparameter tuning.
- Collaborated with researchers to analyze financial market data and enhance investment strategies.

## PROJECTS

---

### Analysis of Tree Distribution and Fire Risk in Los Angeles

May. 2025

- Merged wildfire, tree inventory, and environmental datasets using unique identifiers to create a unified analytical dataset for fire risk assessment.
- Performed data cleaning, spatial joins, and categorical encoding to prepare complex real-world data for modeling.
- Applied multinomial logistic regression to evaluate the impact of tree density and species on fire size categories, interpreting model results with odds ratios
- Visualized geospatial patterns of vegetation and fire occurrence to identify high-risk zones and inform urban planning recommendations.

### Analysis of Heart Disease-Related Dataset

May 2023

- Applied decision tree modeling with entropy-based feature selection to predict heart disease, achieving 75% accuracy.
- Utilized Python (NumPy, Pandas, Scikit-learn) for exploratory data analysis and feature engineering on CDC BRFSS 2015 survey data.
- Implemented hierarchical clustering heatmaps and contingency tables to visualize relationships between health indicators and heart disease.
- Conducted odds ratio analysis to assess the impact of high blood pressure, cholesterol, and age on heart disease prevalence.

### Comparative Analysis of Optimization Techniques for Logistic Regression

Mar. 2023

- Implemented and compared SGD, Mini-Batch, Full Batch Gradient Descent, and Newton's Method for logistic regression optimization in R.
- Designed and executed large-scale simulations to evaluate computational efficiency and predictive accuracy of optimization techniques.
- Applied Cholesky decomposition to enhance Newton's Method efficiency and benchmarked models for performance trade-offs.
- Utilized R and LaTeX for algorithm implementation, data analysis, and professional report formatting.

### Contributions to Type II Diabetes

Nov. 2022 - Dec. 2022

- Extracted large-scale survey data from CA.GOV and the CDC BRFSS using API methods, then cleaned and preprocessed datasets to ensure accuracy and usability for analysis.
- Developed a diabetes prediction model using Random Forest and Logistic Regression, optimizing feature selection and addressing class imbalance to improve model accuracy.
- Applied linear regression and data visualization techniques to investigate correlations between income, demographics, and diabetes prevalence, identifying key risk factors.
- Scraped grocery store data to analyze relationships between food nutrition and pricing, employing linear regression and feature selection techniques to refine insights.

## TECHNICAL SKILLS

---

**Programming:** Python, R, SQL, HTML, CSS, JavaScript

**Tools & Platforms:** MySQL, LaTeX, Microsoft Office, Tableau

**Technical Skills:** Data Analysis, Machine Learning, Data Visualization, Web Development

**Certifications:** ACCA – Passed 5/13 exams (Business and Technology, Management Accounting, Financial Accounting, Performance Management, Financial Reporting)