RUITIAN LIU

Mobile: (+86)13045079645 Email: yktyxtt@outlook.com

EDUCATION

North China Institute of Science and Technology,   Sept 2021 – Jun 2025
Major: Applied Statistics (B.S.)
Relevant Coursework: Multivariate Statistical Analysis (96); Dissertation (95); Sampling Technique (95); Nonparametric Statistics (91); Time Series Analysis (92); Data Mining (91); Categorical Data Analysis (92); Regression Analysis (89); Probability Theory (85); Mathematical Statistics (85).

PROFESSIONAL EXPERIENCES

Breast Cancer Diagnosis via Statistical & Machine Learning Analysis   Winter 2024
Directed by Prof. Meiling Chen

· Analyzed the Wisconsin Diagnostic Breast Cancer dataset using multivariate statistics (PCA, LDA, correlation analysis) and t-SNE visualization to identify key features distinguishing tumors.
· Developed predictive models including LASSO-Logistic Regression (for feature selection/interpretability) and multiple ML algorithms (K-means, Decision Trees, Random Forest, SVM, GBM, AdaBoost, ANN), achieving 99% accuracy with a tuned SVM model.
· Validated model robustness on held-out patient cases and created interpretable radar charts for clinical intervention, visualizing individual patient profiles against benign baselines.

Machine Operation Anomaly Detection   Summer 2025
Funding supported by Geely Automobile Group, directed by Prof. Hua Shang

· Built feature engineering pipeline incorporating Cartesian product encoding, sliding-window statistics, and PrefixSpan for state transition pattern mining.
· Resolved zero-inflated extreme class imbalance through SMOTE under-sampling of normal operations, Isolation Forest, and cost-sensitive modeling of XGBoost.
· Developed hybrid deep learning framework using Transformer encoders & Autoencoders for temporal dependencies and anomaly scoring, combined with BiLSTM-CRF next-step prediction and state transition matrices, achieving 80%+ prediction recall rate of all operation anomaly types.

Nonlinear Impacts of SNS Social Support on Gen Z's Mental Health   Dissertation
Directed by Prof. Wenguang Yang

· Developed a stratified sampling framework with Neyman optimal allocation, reducing sampling error by 18.4% (Deff=0.816) and ensuring demographic representativeness across five-tier Chinese cities.
· Developed LASSO-regularized Spearman Correlation Networks for nonlinear relationship exploration, integrating Spinglass clustering to identify variable communities and centrality metrics to pinpoint key influencers on mental health.
· Utilized Generalized Additive Mixed Models (GAMMs) to model complex nonlinear effects and interaction terms, conducting threshold effect analysis with Bootstrap validation to identify significant inflection points in support-impact relationships.
· Applied Propensity Score Matching (PSM) and testing for mediation effects to address potential confounding and validate model robustness, particularly regarding observed gender differences in mental health outcomes.

COMPETITION EXPERIENCES

Marketing Research on Cultural Education of Beijing for Gen-Z   Group Leader
National Second Prize (hosted by Chinese Business Statistics Society)

· Taught group members with various skills, including EDA visualizations in R, cloudmap drawing in Python, ANOVA, tests of questionnaire, logistic regression, and contingency analysis.
· Designed the entire questionnaire and applied Two-Stage Sampling technique distributing 763 questionnaires to each institution within districts in Beijing with a permissible error of 0.04.
· Employed Structural Equation Model (SEM) to explore potential variables and causal relationships between them, providing a systematic understanding of the complex relationship among respondents' culture preference and behavior, as well as their characteristics and family backgrounds.
· Gained further understanding through Canonical Correlation Analysis studying, which studied variables, including the pairs eliminated by Confirmatory Factor Analysis and GFI-test in SEM.

Logistic Network Optimizations   Group Leader
National Second Prize (hosted by Chinese Society of Optimization, Overall Planning and Economic Mathematics)

· Cleaned outliers, imputed missing values using KNN, and pivoted data to obtain the total cargo volume of 2023 in each venue.
· Utilized Dynamic Time Wrapping algorithm to cluster time series data of cargo volume in each lane, described the feature of each cluster, and predict cargo volume in 2023 using AutoARIMA model.
· Optimized cargo volume fluctuation minimization under constraints (venue closures, capacity limits, lane equilibrium) using Simulated Annealing, achieving an average lane loading rate of 5.29%.

SKILLS

Programming:	R (ggplot2, plotly, dplyr, e1071, caret, Hmisc, TSA, forecast, MASS, h2o, mgcv, glmnet, lme4, rstanarm, np, car, FactoMineR, parallel, broom), Python (keras, scikit-learn, matplotlib, statsmodels, scipy, plotly, dask, seaborn, nltk, pingouin, tensorflow, pytorch, sktime, opencv, tsfresh), Matlab, C, MySQL, Stata
Softwares and Tools:	SPSS, LaTeX, Excel, Amos, Eviews, PowerBI, Tableau