Rui Shi

Auburn, AL 36830

(732) 289-0096   rzs0112@auburn.edu

Professional Summary

Ph.D. candidate in Statistics with expertise in outlier detection, clustering, classification, Bayesian Statistics, and other machine learning methods. Skilled in quantitative models, data science, and financial risk management, with a strong record of interdisciplinary research and teaching.

Publications

1 Rui Shi, Nedret Billor, and Elvan Ceyhan. Outlier Detection with Cluster Catch Digraphs. *ArXiv Preprint ArXiv:2409.11596*, 2024. (under peer review process)

2 Rui Shi, Nedret Billor, and Elvan Ceyhan. Outlyingness Scores with Cluster Catch Digraphs. *ArXiv Preprint ArXiv:2501.05530*, 2025.

3 Rui Shi, Nedret Billor, and Elvan Ceyhan. Cluster Catch Digraphs with Nearest Neighbor Distance. *ArXiv Preprint ArXiv:2501.06268*, 2025.

Academic Presentation

Outlier Detection with Cluster Catch Digraphs (15 minutes), ICORS meets DSSV 2024   *08.2024*
Outlier Detection with Geometric Random Digraphs (6 minutes), SDSS 2024 Conference *06.2024*
Cluster Catch Digraphs with the Nearest Neighbor Distance (45 minutes), Graduate Student Seminar, the Department of Mathematics and Statistics of Auburn University   *11.2023*

Education

Doctor of Philosophy (candidate): Mathematics (Statistical track)
Auburn University, Auburn, AL, USA   *08.2019 - 05.2025 (expected)*

- Dissertation: Outlier Detection with Cluster Catch Digraphs
- Emily Virginia Haynsworth Memorial Endowment for Excellence in Mathematics   *12.2023*
- GPA: 3.88/4.0

Master of Science: Financial Statistics & Risk Management
Rutgers University - New Brunswick, NJ, USA   *08.2017 - 05.2019*

- GPA: 3.95/4.0

Bachelor of Science: Statistics
Sun Yat-sen University, Guangzhou, Guangdong, China   *09.2012 - 06.2016*

Research Projects Experiences

Outlier Detection with Cluster Catch Digraphs (CCDs) (doctoral dissertation)
06.2022-12.2024

- Developed and implemented eight novel CCD-based outlier detection algorithms and outlyingness scores.
- Optimized each proposed algorithm independently with R and Python on Windows, Linux, and high computing clusters
- Achieved exceptional outlier detection accuracy (95% mean accuracy), surpassing established methods like LOF and DBSCAN.
- Composed analytical reports of over 150 pages with Latex for publications.
- Earned Emily Virginia Haynsworth Memorial Endowment for Excellence in Mathematics.

Cluster Catch Digraphs with Nearest Neighbor Distance (doctoral dissertation)
06.2022-12.2024

- Developed UN-CCDs, a novel clustering method utilizing Nearest Neighbor Distance (NND) for spatial randomness test.
- Addressed the limitation of existing CCD methods in high-dimensional data, and improved clustering performance with NND.
- Strengthened the robustness of UN-CCDs by incorporating several enhancements including multiple comparison with Holm's correction, descending radius testing, lower-tailed testing, etc.
- Demonstrated the effectiveness of UN-CCDs with both uniform and Gaussian clusters through extensive simulations, achieved up to 215% improvement in F_2-scores compared to existing CCD methods.

The Erdős Institute Project: Identifying Facial Expression with Machine Learning Models
09.2024 - 12.2024

- Worked on a data set with 36k images consisting of 7 different facial expressions (i.e., happy, disgust, neutral, etc.).
- Implemented progressive ML approaches starting with dimension reduction via Eigenface, followed by traditional classifiers (e.g., KNN, QDA, and Random Forests, etc.), achieved baseline average F-score of 0.54.
- Developed advanced feature extraction pipeline using Histogram of Oriented Gradients (HOG), combining XGBoost for classification, and improving model performance by 11%.
- Designed and tuned CNN (Convolutional Neural Network) architecture with 20 layers, delivered 16% performance improvement and reached an F-score of 0.70, surpassing all previous approaches, beating the best results on Kaggle.
- Earned "Distinction" award (top 10) among 60+ teams consisting of Ph.D. and postdoctoral researchers from top-tier universities.

Identify Material Degradation by Deep Learning Approach
03.2021 - 05.2021

- Identified the degree of the degradation of polymer materials with computer vision techniques and CNN.
- Developed of an automatic monitoring system that alarms when the degradation touches a threshold.

- Performed image analysis and model training with Python using TensorFlow.
- Achieved an overall 92% validation accuracy with 200+ individual photos of polymer materials.

Forest Cover Types Prediction *11.2018 - 12.2018*

- Predicted forest cover type from cartographic variables.
- Conducted and compared supervised methods such as SVM, LDA, KNN, multinomial logistic regression, and unsupervised methods, including clustering and deep learning.
- KNN (K Nearest Neighbors) outperforms other approaches with a 96% overall prediction accuracy in cross validation, which validated the weak model assumption.

Single Stocks Selection in MSCI World Index to Gain from Market Short Squeeze
10.2018

- Investigated short squeeze potentials from more than 160 stocks.
- Extracted over 10000 data sets across different periods when the market fluctuated wildly.
- Applied LDA, QDA, decision trees, and logistic regression on year-long data to model the probabilities and intensities of short squeeze events. Utilized cross-validation on model selection.
- Factored EPS (Earning Per Share), PE (Price Earnings) ratio, volatility, and market capital in time-series regression to catch price movement.
- Achieved an average monthly profit of 20.9% on historical test price movement series (virtual account).

Teaching Experience

Department of Mathematics and Statistics, Auburn University:
MATH 1610-Calculus I, Fall 2024 (recitation leader)
STAT 3611-Probability and Statistics II Laboratory, Summer 2024 (instructor)
STAT 3611-Probability and Statistics II Laboratory, Spring 2024 (instructor)
STAT 7000-Experimental Statistics I, Spring 2024 (TA)
MATH 2630-Calculus III, Fall 2023 (recitation leader)
MATH 1610-Calculus I, Spring 2023 (recitation leader)
MATH 1610-Calculus I, Fall 2022 (recitation leader)
MATH 2630-Calculus III, Spring 2022 (TA)
STAT 2510-Statistics for Biological and Health Sciences, Spring 2022 (TA)
MATH 2630-Calculus III, Fall 2021 (recitation leader)
MATH 1120-Pre-Calculus Algebra, Spring 2021 (instructor)
MATH 1120-Pre-Calculus Algebra, Fall 2020 (instructor)
MATH 2630-Calculus III, Spring 2020 (TA)
STAT 2510-Statistics for Biological and Health Sciences, Fall 2019 (TA)

Skills

Core Domain Expertise: Data Science, Machine Learning, Programming, Quantitative models, Financial Risk Manager, Teaching.
Programming: R, Python, Latex, Linux, SQL, Microsoft Office.

Soft Skills: Leadership and mentorship, efficient team worker, effective communicator.
Other Skills: Effective analytical persona, great time manager, resilient and steadfast, enthusiastic learner.

Work History

Graduate Teaching Assistant
Auburn University, Auburn, AL, USA   *08/2019 - present*

- Led undergraduate recitations and delivered lectures for Calculus I to III and introductory Statistics and Probability courses, ensuring student comprehension of complex mathematical concepts.
- Facilitated interactive learning experiences that promoted student engagement and critical thinking, as evidenced by consistently positive student feedback.

Fixed Income Department Intern
IFS Securities, Atlanta, GA, USA   *02/2019 - 05/2019*

- Developed a platform/framework based on data mining to decide the bid price on municipal bonds.
- Amplified the efficiency of IFS bond market activities in the primary markets.
- Optimized the pricing algorithms and implemented them on the platform to maximize the returns.
- Provided bankers guidance and suggestions to choose targeted issuers and predicted their cash flow needs in the future.

Foreign Exchange Trader Intern
HEZY International Financial Services Co. Ltd., Shenzhen, China   *06/2018 - 08/2018*

- Performed professional methodologies (e.g., moving average system, chaos theory) to predict the trend of popular currency pairs (e.g., XAUUSD, EURUSD, and USDJPY).
- Developed investment strategies using advanced technical indicators and statistical modeling.
- Managed a real account of $2000 with 1:100 leverage and realized a profit of 19.1% in a month.

Grader
Rutgers University - New Brunswick, NJ, USA   *01/2018 - 05/2018*

- Evaluated and provided feedback on undergraduate Statistics course work, ensuring accurate assessment of student understanding and timely return of assignments.
- Completed the grading assignment for over 100 students across an entire semester, maintaining a 48-hour turnaround time and achieving high consistency and accuracy.

Purchasing and Budgeting Department Intern
Yonyou Network Technology Co. Ltd., Beijing, China   *05/2017 - 08/2017*

- Formulated purchasing budgets for construction projects with senior workers, checked the quality of building materials prior to purchase, and uploaded budgets into the company's accounting system.
- Negotiated and verified contract provisions with vendors to ensure clarity and accuracy.
- Successfully managed a budget of over 0.5 million CNY across 4 construction projects, achieving 16% cost savings through strategic and rigorous planning on purchases, contracts, time frames, etc.

Certifications

The Erdős Institute Data Science Boot Camp Certificate   *12.2024*
Financial Risk Manager (FRM) Part I, Global Association of Risk Professionals (GARP) *11.2022*

Awards

Distinguished Data Science Project Award, the Erdős Institute   *12.2024*
COSAM Graduate Student Travel Award   *08.2024*
Graduate Student Council Travel Award, Auburn University   *07.2024*
Emily Virginia Haynsworth Memorial Endowment for Excellence in Mathematics, Auburn University   *12.2023*

Activities

Volunteer of 2024 ASA Alabama-Mississippi Chapter Annual Conference and Business Meeting
03.2024
Organizer for the 11th & 12th English Speaking Contest of Sun Yat-sen University   *2013 - 2014*