conferenceseries.com

10ᵗʰ International Conference on DATA SCIENCE AND MACHINE LEARNING APPLICATIONS

August 08, 2022 | Webinar

Stacked Generalizations in Imbalanced Fraud Data Sets using Resampling Methods

Kathleen Kerwin
Department of Defense (CAMS MSDS CFE MBA)

Statement of Problem: Predicting fraud is challenging due to inherent issues in the fraud data structure since the crimes are committed through trickery or deceit with an ever-present moving target of changing modus operandi to circumvent human and system controls. As a national security challenge, criminals continually exploit the electronic financial system to defraud consumers and businesses by finding weaknesses in the system, including in audit controls. The purpose of this study is to identify fraudulent transactions in imbalanced datasets. Methodology and theoretical orientation: This study applies stacked generalization using meta or super learners for improving the performance of algorithms in step one (minimizing the algorithm error rate to reduce its bias in the learning set) and then in step two the results are input into the meta learner with its stacked blended output (with the weakest algorithms learning better). A fundamental key to fraud data is that it is inherently not systematic, and an optimal resampling methodology has yet not been identified. Findings: Building a test harness, for all permutations of algorithm sample set pairs, demonstrates that the complex, intrinsic data structures are all thoroughly tested. Conclusion and significance: A comparative analysis on fraud data that applies stacked generalizations provides useful insight to find the optimal mathematical formula for imbalanced fraud data sets necessary to improve upon fraud detection for national security.

Biography

Kathleen is the corresponding author. She has had a diverse background in law enforcement, as well as large application and system software development and management working for a top tiered bank and two Big 5 Consulting firms. She passed the CPA exam and is a Certified Fraud Examiner, as well as a Certified Anti Money Laundering Specialist. Kathleen earned her Master's in Data Science Program at Northwestern where she completed her thesis in credit risk scoring. This paper is an aggregation of a lifetime of education, practical work experience and research. Currently Kathleen is working as a Data Scientist for the Department of Defense.

DATASCIENCE CONFERENCE 2022
August 08, 2022 Webinar

Journal of Information Technology &
Software Engineering

Volume: 12

Page 1