



You are invited to attend!

Medicare Fraud Detection using Machine Learning Models

A Mathematics Master's Project Presentation

By Marcel Tiono

Department of Mathematics

University of Nebraska at Omaha

Please join us on Tuesday, July 25th, at 9 am via zoom

Zoom link: <https://unomaha.zoom.us/j/97904603038>

Academic Advisor: Dr. Xiaoyue Cheng, Associate Professor

Department of Mathematical and Statistical Sciences, UNO

External Advisor: Tyler Billings, Data Scientist

Aligned Assurance Operations, Mutual of Omaha

Abstract

Medicare is a United States government healthcare program that provides affordable healthcare insurance to elderly over 65 years old and younger with disabilities. The demand for Medicare insurance is increasing due to the increasingly aging population and creates an environment with increased risk of fraud. To fight fraudulent activities, machine learning approaches have been introduced in Medicare to identify fraudulent behavior and bad actors. This study evaluated previous fraud detection methods, which leveraged supervised machine learning models to detect Medicare fraud, using Logistic Regression (LR), XGBoost, and Light Gradient Boosting models (LGBM). Two datasets were used to perform the study: Medicare's Physicians & Other Practitioners dataset and the Lists of Excluded Individuals/Entities (LEIE) dataset by the Office of the Inspector General. With the LEIE dataset, we mapped the Medicare data using the National Provider Identifier (NPI) and year. The resulting data is highly imbalanced. Then three resampling methods ROS and RUS, ROSE, SMOTE were tested to address the problem of class imbalance data. The results show that XGBoost performs better than LGBM and LR learners with an average Area Under Curve (AUC) score of 0.9632 and average True Positive Rate (TPR) of 0.9677 when using ROS and RUS technique. LR performs better when using ROSE technique with an average AUC score of 0.9273. Based on the performance, we can say that XGBoost model used with ROS- RUS is better in identifying Medicare fraud committed by the healthcare provider than LGBM and LR with the Medicare data sample used in this study.