

Analysis of Data Balancing and Advanced Machine Learning Techniques

Mr. P. Manikanda Prabu¹, R. Dharaniga², S. Kuralarasi³, S. Meenalochani⁴

Professor, Department of Computer Science and Engineering¹

Student, Department of Computer Science and Engineering^{2,3,4}

Anjalai Ammal Mahalingam Engineering College, Thiruvavur, Tamilnadu, India

maniadt2006@gmail.com, dharanigaravi2004@gmail.com, kuralarasics5@gmail.com,

meenalochnisakthivel@gmail.com

Abstract: Medical insurance fraud poses a gigantic threat to healthcare systems globally with enormous financial losses and discrediting legitimate claimants. Below is the implementation of a machine learning model for medical insurance fraud detection with focus on addressing class imbalance using Synthetic Minority Over-sampling Technique (SMOTE). The system suggested here is having multiple modules like data gathering, preprocessing, feature extraction, model training, and real-time fraud detection. Various machine learning algorithms like Decision Trees and XGBoost were compared on the efficiency level to identify sophisticated patterns of fraud. The SMOTE technique was utilized to balance the data, and there was significant improvement in model performance by generating synthetic samples for the minority (fraud) class. The model was trained and validated on actual healthcare claim data, and performance was measured in terms of accuracy, precision, recall, and F1-score. Results indicate improved detection rates and reduced false positives compared to traditional models. This implementation is designed to help healthcare providers and payers minimize financial risk and improve claims handling integrity using a clever, scalable, and comprehensible solution for fraud detection.

Keywords: Medical Insurance Fraud, Machine Learning, SMOTE, Fraud Detection, XGBoost

