INTERNATIONAL JOURNAL OF CURRENT ADVANCED RESEARCH
ISSN: O: 2319-6475, ISSN: P: 2319-6505, Impact Factor: 6.614
Available Online at www.journalijcar.org
Volume 13; Issue 12; December 2024; Page No.3396-3401
DOI: http://dx.doi.org/10.24327/ijcar.2024.3401.1722

Research Article

ENHANCING MEDICARE FRAUD DETECTION THROUGH ML: ADDRESSING CLASS IMBALANCE WITH SMOTE-ENN

O. Ramyateja[1], CH. Deekshitha[2], B. Kavya Sri[3], Fatima Tabassum[4]
Assistant Professor[1], UG Student [2,3,4]
Department of it, Malla Reddy Engineering College for Women (UGC-Autonomous),
Maisammguda, Hyderabad, Telangana-500100

ARTICLE INFO

Article History:

Received 20th October, 2024
Received in revised form 10th November, 2024
Accepted 20th November, 2024
Published online 28th December, 2024

Key words:

Healthcare fraud, imbalanced data, machine learning (ML), noisy data.

ABSTRACT

he healthcare fraud detection field is constantly evolving and faces significant challenges, particularly when addressing imbalanced data issues. Previous studies mainly focused on traditional machine learning (ML) techniques, often struggling with imbalanced data. This problem arises in various aspects. It includes the risk of overfitting with Random Oversampling (ROS), noise introduction by the Synthetic Minority Oversampling Technique (SMOTE), and potential crucial information loss with Random Undersampling (RUS). Moreover, improving model performance, exploring hybrid resampling techniques, and enhancing evaluation metrics are crucial for achieving higher accuracy with imbalanced datasets. In this paper, we present a novel approach to tackle the issue of imbalanced datasets in healthcare fraud detection, with a specific focus on the Medicare Part B dataset. First, we carefully extract the categorical feature ''Provider Type'' from the dataset. This allows us to generate new, synthetic instances by randomly replicating existing types, thereby increasing the diversity within the minority class. Then, we apply a hybrid resampling method named SMOTE-ENN, which combines the Synthetic Minority Oversampling Technique (SMOTE) with Edited Nearest Neighbors (ENN). This method aims to balance the dataset by generating synthetic samples and removing noisy data to improve the accuracy of the models. We use six machine learning (ML) models to categorize the instances. When evaluating performance, we rely on common metrics like accuracy, F1 score, recall, precision, and the AUC-ROC curve. We highlight the significance of the Area Under the Precision-Recall Curve (AUPRC) for assessing performance in imbalanced dataset scenarios. The experiments show that Decision Trees (DT) outperformed all the classifiers, achieving a score of 0.99 across all metrics.

INTRODUCTION

Healthcare systems globally face a significant challenge due to fraud, which impacts both their financial stability and moral principles. In particular, the U.S. Medicare program, a key element of the healthcare sector, experiences substantialfinancial loss from such fraudulent practices. According to the Federal Bureau of Investigation, healthcare fraud represents 3–10% of the total healthcare costs, leading to yearly losses between $19 billion and $65 billion [1]. These illegal activities not only deplete financial resources but also affect the operational efficiency and trustworthiness of healthcare systems. Therefore, it is imperative to implement effective and strong fraud detection strategies, especially in Medicare, which serves a broad and diverse population. Ensuring efficient fraud detection is vital for the protection of public funds and guaranteeing that resources are distributed fairly for necessary healthcare services and patient care. The challenge in healthcare fraud detection lies in the evolving nature of fraud schemes, which are complex and diverse. Traditional, rule-based detection methods fall short in this dynamic environment, lacking the necessary adaptability andscalability to address the sophisticated nature of modern healthcare fraud. Machine learning (ML), a subfield o Artificial intelligence (AI) has demonstrated exceptionalproficiency in healthcare fraud detection, particularly in processing the Medicare dataset released annually by the U.S. government [2]. This dataset is a crucial resource for researchers focusing on healthcare fraud detection. This reflects the government's

*Corresponding author: **O. Ramyateja**

Department of it, Malla Reddy Engineering College for Women (UGC-Autonomous), Maisammguda, Hyderabad, Telangana-500100

commitment to combating fraud by equipping specialists with vital data, thereby facilitating the development of more sophisticated fraud detection strategies based on ML. Its strength lies in its ability to learn from historical data and adapt to emerging fraudulent patterns, making it effective in analyzing large datasets to identify anomalies and fraud indicators. This adaptability renders ML indispensable in creating efficient, responsive systems for large-scale operations like Medicare, positioning it is an indispensable asset in combating healthcare fraud [3]. Machine Learning (ML), however, excels in this aspect. Its ability to learn from historical data and adjust to new fraudulent patterns allows it to process and analyze vast datasets, detecting anomalies and patterns indicative of fraud. This capability positions ML as a crucial tool in creating more effective and responsive fraud detection systems, especially for large-scale operations like Medicare. Its dynamic approach makes it an indispensable asset in the ongoing fight against healthcare fraud [3]. Recent studies, such as those by [4], [5], [6], [7], and [8] demonstrate the successful application of ML techniques using the Medicare dataset to uncover fraudulent activities. The Medicare datasets [9], disseminated by the Centers for Medicare and Medicaid Services, exhibit a pronounced class imbalance characterized by a disproportionate representation of non-fraudulent cases relative to fraudulent instances. This class imbalance presents a formidable impediment to the efficacy of ML algorithms deployed for fraud detection. Predominantly, ML models are predisposed to a bias towards the majority class, in this case, non-fraudulent transactions, leading to a heightened incidence of false negatives. This phenomenon occurs when the algorithm erroneously categorizes fraudulent activities as legitimate, a direct consequence of the skewed training data [2], [10]. Such imbalance in the dataset precipitates the development of ML models that demonstrate suboptimal performance in the accurate detection of fraudulent activities. This deficiency critically undermines the overarching effectiveness and reliability of the fraud detection mechanism within the healthcare domain. To ameliorate this situation, it is imperative to establish datasets that are balanced, thereby ensuring that ML algorithms are more adept at discerning the minority class, which in this context refers to fraudulent transactions. A balanced dataset is instrumental in enabling the algorithm to detect nuanced patterns and anomalies that are indicative of fraudulent activities [5]. A notable gap in current research endeavors within healthcare fraud detection is the inadequate focus on addressing the challenges posed by imbalanced data. The preponderance of research has been directed towards classification tasks, with insufficient attention to the intricate issue of data imbalance. Although there has been a notable deficiency in addressing data imbalances within healthcare fraud detection, some researchers have begun to address this gap using resampling techniques. These methodologies, which include Random Oversampling (ROS) [5], Adaptive synthetic sampling approach (ADASYN) [11], and Synthetic Minority Over-sampling Technique (SMOTE) [12]. Concurrently, undersampling of the majority class is executed using Random Undersampling (RUS) [13] to achieve a balanced dataset. Despite the efficacy of these techniques, challenges persist. ROS methods, for instance, may be susceptible to overfitting, potentially compromising the generalizability of the model. Meanwhile, the application of SMOTE carries the risk of introducing noise to the dataset. Moreover, the implementation of RUS comes with its own set of concerns, notably the risk of discarding crucial data, potentially leading to a loss of important information. The intricate trade-offs andconsiderations associated with each resampling technique underscore the complexity of addressing the class imbalance in healthcare fraud detection datasets. To address the limitations identified in prior research, we focus on three main areas:

- Advancing research into techniques for managing imbalanced datasets
- Evaluating resampling approaches, with an emphasis on the drawbacks of ROS, which can cause overfitting, and SMOTE, which may add noise to the dataset.
- Examining the impact of RUS on the potential loss of essential data, which could lead to overlooking critical indicators of fraud. This paper introduces a novel approach to address imbalanced datasets in healthcare fraud detection, particularly focusing on the Medicare Part B dataset. A key innovation lies in the meticulous separation of the categorical features from the numerical features, enabling the random generation of synthetic instances to enrich minority class diversity. Our proposed Synthetic Minority Over-sampling technique with Edited Nearest Neighbors (SMOTE-ENN) hybrid resampling method contributes significantly by simultaneously rebalancing the dataset and eliminating noisy data, which is thenevaluated using various ensemble classifiers. To the best of our knowledge, this paper proposes an approach that combines the separate generation of categorical features, withthe SMOTE-ENN technique and a variety of ensemble learning classifiers. Additionally, we incorporate the use of the Area Under the Precision-Recall Curve (AUPRC) metric for evaluation, enhancing the robustness and comprehensivenessof our analysis.

The main contributions of this paper can be summarized as follows:

- Randomly generate the categorical feature ‘‘Provider. Type’’ based on existing categories in the dataset
- Application of the SMOTE-ENN hybrid resampling method to balance the dataset and remove noisy data.
- Evaluation of the effectiveness of the proposed approach using ensemble learning classifiers.
- Employing the Area Under the Precision-Recall Curve (AUPRC) metric for a more effective evaluation of model performance in the context of an imbalanced dataset

LITERATURE REVIEW

1. R. Bauder and T. Khoshgoftaar, “Medicare fraud detection using random forest with class imbalanced big data,” in Proc. IEEE Int. Conf. Inf. Reuse Integr. (IRI), Jul. 2018, pp. 80–87.

The paper addresses the challenge of detecting Medicare fraud using large, imbalanced datasets. Medicare fraud is a significant issue in healthcare, where fraudulent activities can lead to substantial financial losses. Traditional detection

methods often struggle with the class imbalance problem, where fraudulent instances are rare compared to legitimate transactions, making it difficult for classifiers to effectively identify fraudulent cases. The authors propose using the Random Forest (RF) algorithm, a powerful machine learning technique, to detect fraud within this imbalanced dataset. They explore the challenges posed by class imbalance and discuss several techniques to handle this issue, such as data resampling, cost-sensitive learning, and ensemble methods. The paper also emphasizes the importance of feature selection in building an effective fraud detection model, where they identify key features that are most indicative of fraudulent activities. Through experimentation, the authors demonstrate that Random Forest can significantly improve the detection of Medicare fraud, even in the presence of class imbalance, by incorporating these techniques. The results highlight the model's ability to handle large-scale datasets and imbalances while maintaining a high level of accuracy. The paper concludes by suggesting that Random Forest, combined with appropriate handling of class imbalance, offers a promising approach for fraud detection in healthcare and other domains characterized by imbalanced big data.

2. J. Hancock and T. M. Khoshgoftaar, "Medicare fraud detection using CatBoost," in Proc. IEEE 21st Int. Conf. Inf. Reuse Integr. Data Sci. (IRI), Aug. 2020, pp. 97–103.

The paper explores the application of CatBoost, a gradient boosting algorithm, to detect Medicare fraud. Medicare fraud detection is a challenging problem due to the imbalanced nature of the data, where fraudulent instances are rare compared to legitimate transactions. The authors focus on leveraging CatBoost, an advanced machine learning algorithm known for its efficiency and robustness, particularly in handling categorical features and imbalanced datasets. The paper highlights the advantages of CatBoost over traditional machine learning models, such as decision trees and random forests, especially in the context of large-scale healthcare datasets. Through their experiments, the authors demonstrate that CatBoost effectively identifies key patterns in Medicare claims data, providing an accurate and scalable solution for fraud detection. The study also discusses the importance of feature engineering, where the authors identify critical features that can help distinguish between legitimate and fraudulent claims. The results indicate that CatBoost outperforms several other algorithms in terms of accuracy, precision, and recall, making it a strong candidate for detecting fraud in healthcare. Furthermore, the paper discusses the practical challenges in deploying machine learning models in real-world fraud detection systems and suggests strategies for model optimization and evaluation. The authors conclude that CatBoost is a promising approach for Medicare fraud detection, offering significant potential for improving fraud detection systems in healthcare and other sectors facing similar challenges.

3. M. Herland, T. M. Khoshgoftaar, and R. A. Bauder, "Big data fraud detection using multiple medicare data sources," J. Big Data, vol. 5, no. 1, pp. 1–21, Dec. 2018.

The paper explores the use of big data techniques for detecting fraud in Medicare claims by leveraging multiple data sources. Medicare fraud detection remains a complex problem due to the large volume and variety of healthcare data, as well as the imbalanced nature of fraudulent claims compared to legitimate ones. The authors propose an integrated approach that utilizes data from multiple sources, such as billing data, patient records, and healthcare provider information, to improve fraud detection accuracy. By combining diverse data streams, the method enhances the model's ability to detect fraudulent activity by capturing a broader range of patterns and anomalies. The paper emphasizes the importance of data preprocessing, feature selection, and feature engineering in dealing with large, high-dimensional datasets, and proposes several techniques to handle these challenges effectively. The authors experiment with different machine learning algorithms, demonstrating the benefits of using ensemble methods and hybrid models to detect fraud more efficiently. The results show that integrating multiple data sources significantly improves detection performance compared to using a single data source. Furthermore, the study highlights the need for scalability and robustness in fraud detection systems, given the constantly growing volume of Medicare data. The authors conclude that leveraging big data and machine learning techniques across multiple data sources is an effective strategy for improving fraud detection in Medicare and can be extended to other domains with similar data challenges.

4. M. Rashid, J. Kamruzzaman, T. Imam, S. Wibowo, and S. Gordon, "A tree-based stacking ensemble technique with feature selection for network intrusion detection," Appl. Intell., vol. 52, no. 9, pp. 9768–9781, 2022.

The paper presents a novel approach for improving network intrusion detection using a tree-based stacking ensemble method combined with feature selection. Network intrusion detection is a critical aspect of cybersecurity, as it involves identifying unauthorized or malicious activity in a network. The challenge lies in accurately detecting such intrusions amidst large and complex datasets, often containing irrelevant or redundant features. To address this, the authors propose a tree-based stacking ensemble model that combines multiple decision tree-based classifiers to improve the accuracy and robustness of intrusion detection systems. The stacking ensemble method utilizes a meta-learner to integrate the predictions of base models, enhancing overall performance. Additionally, the paper incorporates a feature selection step to reduce dimensionality and remove irrelevant features, improving model efficiency and reducing overfitting. The authors demonstrate that this hybrid technique outperforms traditional machine learning models in terms of accuracy, precision, and recall. Experimental results show that the proposed method effectively detects a wide range of network intrusions while maintaining computational efficiency. The paper emphasizes the importance of feature selection in enhancing the performance of intrusion detection systems and highlights the advantages of stacking ensembles in improving predictive accuracy. The authors conclude that the tree-based stacking ensemble with feature selection offers a promising approach to network intrusion detection, providing a scalable and effective solution for real-time cybersecurity applications.

5. G. E. A. P. A. Batista, R. C. Prati, and M. C. Monard, "A study of the behavior of several methods for balancing machine learning training data," ACM SIGKDD Explorations Newslett., vol. 6, no. 1, pp. 20–29, Jun. 2004

The paper investigates various techniques for addressing the class imbalance problem in machine learning. Class imbalance occurs when one class in a dataset significantly outnumbers the other, leading to biased models that favor the majority class. This imbalance can degrade the performance of machine learning algorithms, particularly in tasks like fraud detection, medical diagnosis, and intrusion detection, where minority class instances are often more important but harder to predict. The paper provides an in-depth analysis of different methods to balance training data, including oversampling, undersampling, and synthetic data generation techniques. The authors evaluate the effectiveness of these methods on several datasets, comparing their impact on model accuracy, sensitivity, and other performance metrics. The study reveals that different balancing techniques exhibit varying behaviors depending on the nature of the data and the learning algorithm used. Some methods, like random oversampling and SMOTE (Synthetic Minority Over-sampling Technique), tend to improve classification performance, while others, such as random undersampling, may lead to information loss. The paper concludes by emphasizing the importance of selecting the appropriate balancing method for a given problem and dataset, suggesting that no single technique works universally. It calls for further research into hybrid approaches that combine the strengths of multiple techniques to achieve better results in imbalanced data scenarios.

EXISTING SYSTEM

Recent advancements in AI, especially ML, have led to diverse and innovative approaches to detecting healthcare fraud. The authors in [16] aimed to improve decision-treebased ensemble techniques for healthcare fraud detection, utilizing the large Part D Medicare dataset with around 175 million records. The authors in [17] introduced a ML framework that transforms prescription claims into statistical modeling features, focusing on business heuristics, provider prescriber relationships, and client demographics. The study by [2] employed an ensemble feature selection technique in ML models for Medicare fraud detection. This approach improved explainability and reduced data complexity. The work proposed by [18], introduced a Bayesian Belief Network (BBN) model for healthcare fraud detection, involving preprocessing and feature engineering of Texas Medicaid prescription claims. This approach outperformed baseline models in scalability and interpretability. In [19], the authors concentrated on applying a data-centric AI approach to detect U.S. Medicare fraud. This significantly enhanced ML models' performance through careful data preparation and feature engineering. Their approach showed superior results compared to traditional datasets in Medicare fraud classification tasks. Reference [6] proposed a study to detect healthcare fraud instances by applying four ML algorithms. In their research, they identified 19 essential features, which they organized into four primary categories. Upon examining the studies, we can observe the use of diverse methods in detecting fraud, such as ensemble methods, decision-tree-based techniques, and BBN. Moreover, several works emphasize the important role of data preparation, feature engineering, and feature selection in enhancing the model's performance. However, a common limitation observed is the reliance on the significantly imbalanced Medicare dataset for experimentation, an issue that remains largely unaddressed and could potentially result in misclassification outcomes. The paper [20] tackled the problem of imbalanced data by experimenting with different class distributions in their ML models. Using the Medicare Part B dataset, the authors applied six ML models across seven class distributions to address the data imbalance. The results indicate that employing a 90:10 ratio of non-fraud to fraud cases outperformed other models. In their study, [21], the authors addressed the challenge of the imbalanced data in the Medicare dataset by employing ML models for classification and six sampling techniques to balance the dataset. The study's findings demonstrated that RUS consistently gave strong results across all ML models. A semantic embedding approach was proposed in [22]. The author proposed a semantic embedding approach to convert healthcare procedure codes (HCPCS) from the Medicare fraud dataset into semantic embeddings. To address the imbalanced data issue, the work employed a simple undersampling method. Another semantic embedding approach was proposed in [23]. The authors developed semantic embeddings for medical provider types using both pre-trained (Global Vectors for Word Representation (GloVe), Medical Word2Vec (Med-W2V)) and custom (HcpcsVec, RxVec) embeddings from Medicare claims data. This method improved the representation of provider specialties and was validated using various ML algorithms. Additionally, the study tackled the issue of imbalanced data by employing random over-sampling (ROS) and under-sampling techniques. In their study, [24], the authors proposed unsupervised DL techniques to detect procedure code overutilization in medical claims. To tackle the imbalanced data, the test set was composed of outliers representing potential fraudulent cases. The paper, [25], focused on assessing the performance of ML classifiers in the Medicare imbalanced dataset. The authors applied the RUS method with various ensemble learning techniques to address class imbalances. Another paper, [26], explored the classification of healthcare fraud using the highly imbalanced Medicare dataset by employing the RUS method to address the imbalance issue. The results show RUS enhanced the AUC scores while reducing the training data size. In the paper [11], the authors proposed the use of two data balancing techniques, namely: Class Weighing Scheme (CWS) and ADASYN. Moreover, to classify instances, the authors applied a range of ML algorithms.

Disadvantages

- Advancing research into techniques for managing imbalanced datasets
- Evaluating resampling approaches, with an emphasis on the drawbacks of ROS, this can cause overfitting, and SMOTE, which may add noise to the dataset.
- Examining the impact of RUS on the potential loss of essential data, which could lead to overlooking critical indicators of fraud.

PROPOSED SYSTEM

This paper introduces a novel approach to address imbalanced datasets in healthcare fraud detection, particularly focusing on the Medicare Part B dataset. A key innovation lies in the meticulous separation of the categorical features from the numerical features, enabling the random generation of

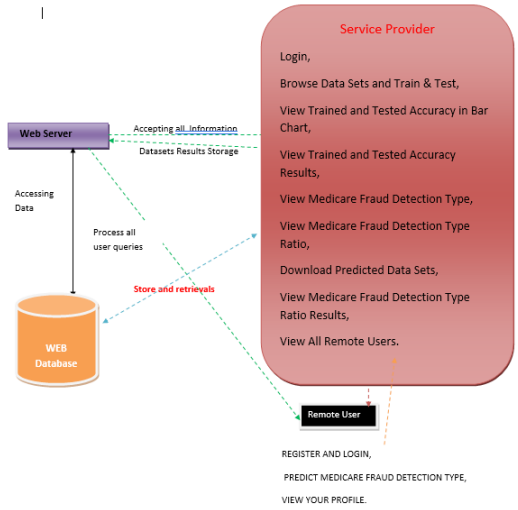
synthetic instances to enrich minority class diversity. Our proposed Synthetic Minority Over-sampling technique with Edited Nearest Neighbors (SMOTE-ENN) hybrid resampling method contributes significantly by simultaneously rebalancing the dataset and eliminating noisy data, which is then evaluated using various ensemble classifiers. To the best of our knowledge, this paper proposes an approach that combines the separate generation of categorical features, with the SMOTE-ENN technique and a variety of ensemble learning classifiers. Additionally, we incorporate the use of the Area Under the Precision-Recall Curve (AUPRC) metric for evaluation, enhancing the robustness and comprehensiveness of our analysis.

Advantages

- Randomly generate the categorical feature ‘‘Provider. Type’’ based on existing categories in the dataset
- Application of the SMOTE-ENN hybrid resampling method to balance the dataset and remove noisy data.
- Evaluation of the effectiveness of the proposed approach using ensemble learning classifiers.
- Employing the Area Under the Precision-Recall Curve (AUPRC) metric for a more effective evaluation of model performance in the context of an imbalanced dataset.

IMPLEMENTATION

SYSTEM ARCHITECTURE



MODULES

SERVICE PROVIDER

In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as Browse Data Sets and Train & Test, View Trained and Tested Accuracy in Bar Chart, View Trained and Tested Accuracy Results, View Prediction Of Botnet Attack Type, View Botnet Attack Prediction Type Ratio, Download Predicted Data Sets, View Botnet Attack Type Prediction Ratio Results, View All Remote Users.

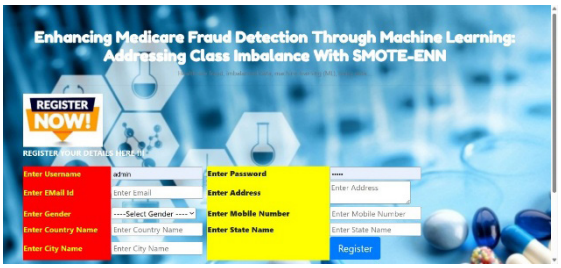
VIEW AND AUTHORIZE USERS

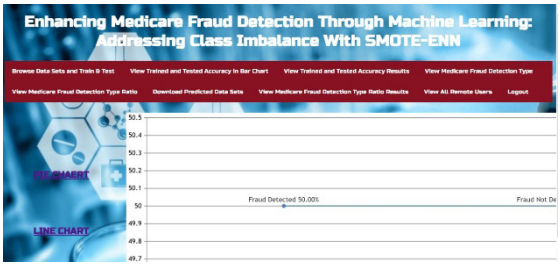
In this module, the admin can view the list of users who all registered. In this, the admin can view the user’s details such as, user name, email, address and admin authorizes the users.

REMOTE USER

In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like REGISTER AND LOGIN, PREDICT BOTNET ATTACK TYPE, VIEW YOUR PROFILE.

RESULT





CONCLUSION

This study emphasizes the need to address imbalanced data in healthcare fraud detection by introducing a novel ML framework based on the SMOTE-ENN hybrid resampling method. This method effectively balances datasets by creating synthetic samples while eliminating noisy data, thereby enhancing the model’s accuracy. Another aspect of our study is the application of the AUC and AUPRC as evaluation metrics. These metrics facilitated a thorough analysis of the models’

performance, with the AUPRC proving to be especially critical in the context of imbalanced datasets. Thus, this approach serves as a basis for new researchers to applynew approaches to detect healthcare fraud. Future research directions include evaluating SMOTE-ENN's performance in diverse healthcare fraud scenarios and combining it with innovative AI technologies such as deep learning (DL) to enhance the effectiveness of fraud detection methods.

References

1. L. Morris, ''Combating fraud in health care: An essential component of any cost containment strategy,'' Health Affairs, vol. 28, no. 5, pp. 1351–1356, Sep. 2009.
2. J. T. Hancock, R. A. Bauder, H. Wang, and T. M. Khoshgoftaar, ''Explainable machine learning models for medicare fraud detection,'' J. Big Data, vol. 10, no. 1, p. 154, Oct. 2023.
3. A. Alanazi, ''Using machine learning for healthcare challenges and opportunities,'' Informat. Med. Unlocked, vol. 30, 2022, Art. no. 100924.
4. R. A. Bauder and T. M. Khoshgoftaar, ''The detection of medicare fraud using machine learning methods with excluded provider labels,'' in Proc. Thirty-First Int. Flairs Conf., 2018, pp. 1–6.
5. R. A. Bauder and T. M. Khoshgoftaar, ''Medicare fraud detection using machine learning methods,'' in Proc. 16th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA), Dec. 2017, pp. 858–865. [Online]. Available: http://ieeexplore.ieee.org/document/8260744/
6. V. Nalluri, J.-R. Chang, L.-S. Chen, and J.-C. Chen, ''Building prediction models and discovering important factors of health insurance fraud using machine learning methods,'' J. Ambient Intell. Humanized Comput., vol. 14, no. 7, pp. 9607–9619, Jul. 2023.
7. P. Dua and S. Bais, ''Supervised learning methods for fraud detection in healthcare insurance,'' in Machine Learning in Healthcare Informatics (Intelligent Systems Reference Library), vol. 56, S. Dua, U. Acharya, and P. Dua, Eds. Berlin, Germany: Springer, 2014, doi: 10.1007/978-3-642- 40017-9_12.
8. R. Bauder, R. da Rosa, and T. Khoshgoftaar, ''Identifying medicare provider fraud with unsupervised machine learning,'' in Proc. IEEE Int. Conf. Inf. Reuse Integr. (IRI), Jul. 2018, pp. 285–292.
9. Centers for Medicare and Medicaid Services. (2017). Research, Statistics, Data, and Systems. [Online]. Available: https://www.cms.gov/researchstatistics-data-and-systems/research-statistics-data-and-systems.html
10. P. Brennan, ''A comprehensive survey of methods for overcoming the class imbalance problem in fraud detection,'' Inst. Technol. Blanchardstown Dublin, Dublin, Ireland, Tech. Rep., 2012.
11. N. Agrawal and S. Panigrahi, ''A comparative analysis of fraud detection in healthcare using data balancing & machine learning techniques,'' in Proc. Int. Conf. Commun., Circuits, Syst. (IC3S), May 2023, pp. 1–4.
12. M. Herland, R. A. Bauder, and T. M. Khoshgoftaar, ''The effects of class rarity on the evaluation of supervised healthcare fraud detection models,'' J. Big Data, vol. 6, no. 1, pp. 1–33, Dec. 2019.

How to cite this article:

O. Ramyateja., CH. Deekshitha., B. Kavya Sri., Fatima Tabassum. (2024) Enhancing medicare fraud detection through ml: addressing class imbalance with smote-enn, *International Journal of Current Advanced Research*, 13(12), pp.3396-3401.

****