Automated Detection of Fraudulent Medicare Providers: ML-Driven Approach for Enhanced Accuracy

M. Venkatesh[1*], A. Sujith Kumar[1], Md. Mansoor[1], G. Satish Chary[1], V. Sai Krishna[1]

[1]Department of Computer Science and Engineering (Data Science), Sree Dattha Group of Institutions, Hyderabad, Telangana, India

[*]Corresponding E-mail: mvenkateshatm1030@sreedattha.ac.in

Abstract

With the overall increase in the elderly population come additional, necessary medical needs and costs. Medicare is a U.S. healthcare program that provides insurance, primarily to individuals 65 years or older, to offload some of the financial burden associated with medical care. Even so, healthcare costs are high and continue to increase. Fraud is a major contributor to these inflating healthcare expenses. The most common method for undertaking the latter involves manually auditing claims data, which is a time-consuming and expensive process. Machine learning models can greatly cut auditing costs by automatically screening incoming claims and flagging up those that are deemed to be suspicious – i.e., potentially incorrect – for subsequent manual auditing. This work provides a comprehensive study leveraging machine learning methods to detect fraudulent Medicare providers. This work uses publicly available Medicare data and provider exclusions for fraud labels to build and assess three different learners. To lessen the impact of class imbalance, given so few actual fraud labels, this framework employs Logistic Regression creating two class distributions. Results show that the other algorithms have poor performance compared with Logistic Regression. Learners have the best fraud detection performance, particularly for the 80:20 class distributions with average AUC scores, respectively, and low false negative rates. This work successfully demonstrates the efficacy of employing machine learning models to detect Medicare fraud.

Keywords: Medical Data, Fraudulent Medicare Provider, Internet of Medical Things, Machine Learning, Data analytics.

1. Introduction

Health insurers receive millions of claims per year. Given that information asymmetries between the principal (insurer) and the agents (health care providers and the insured) can lead to moral hazard, insurance companies face the choice of either paying out insurance claims immediately without any adjustments or reviewing claims that are suspicious. The most common method for undertaking the latter involves manually auditing claims data, which is a time-consuming and expensive process. Machine learning models can greatly cut auditing costs by automatically screening incoming claims and flagging up those that are deemed to be suspicious – i.e., potentially incorrect – for subsequent manual auditing.

Insurance fraud is a widespread and high-priced problem for each policyholder and insurance businesses in all sectors of the coverage industry [1]. India is one of the quickest developing economies in the international, has a burgeoning middle class, and has witnessed a giant upward push within the demand for medical insurance products [2]. Over the last 10 years, the medical health insurance industry has grown at a capital annual compounded boom rate of around 20%. But, with the exponential growth inside the industry, there has additionally been an extended prevalence of frauds within the us. Health insurance fraud contains a huge range of illicit practices and unlawful acts concerning intentional deception or misrepresentation. Statistics mining has an extraordinary effect in enhancing healthcare fraud detection system. Statistics mining has been implemented to fraud detection in both the way i.e., Supervised, and non-supervised way. Information mining strategies and its software for fraud detection in fitness care zone is defined below. In latest years, systems for processing digital claims were increasingly carried out to mechanically perform audits and reviews of claims information. These systems are designed for figuring out regions requiring unique interest together with faulty or incomplete data entry, duplicate claims, and medically non-blanketed services [3]. Even though these structures may be used to locate sure varieties of fraud, their fraud detection competencies are typically restrained because detection particularly is predicated on pre-defined easy guidelines special via domain professionals.

Provider Fraud is one of the biggest problems facing Medicare. According to the government, the total Medicare spending increased exponentially due to frauds in Medicare claims. Healthcare fraud is an organized crime which involves peers of providers, physicians, beneficiaries acting together to make fraud claims. Rigorous analysis of Medicare data has yielded many physicians who indulge in fraud. They adopt ways in which an ambiguous diagnosis code is used to adopt costliest procedures and drugs. Insurance companies are the most vulnerable institutions impacted due to these bad practices. Due to this reason, insurance companies increased their insurance premiums and as result healthcare is becoming costly matter day by day. Healthcare fraud and abuse take many forms.

2. Literature Survey

Herland et. al [4] employed an approach to predict a physician's expected specialty based on the type and number of procedures performed. From this approach, they generate a baseline model, comparing Logistic Regression and Multinomial Naive Bayes, to test and assess several new approaches to improve the detection of U.S. Medicare Part B provider fraud. These results indicate that this proposed improvement strategies (specialty grouping, class removal, and class isolation), applied to different medical specialties, have mixed results over the selected Logistic Regression baseline model's fraud detection performance. Through this work, they demonstrate that improvements to current detection methods can be effective in identifying potential fraud.

Hancock et. al [5] conducted experiments with three Big Data Medicare Insurance Claims datasets. The experiments are exercises in Medicare fraud detection. They show that for each dataset, they obtain better performance from LightGBM and CatBoost classifiers with tuned hyperparameters. Since some features of the data, they are working with are high cardinality categorical features, they have an opportunity to try different encoding techniques in these experiments. They find that across the different encoding techniques, hyperparameter tuning Provides an improvement in the performance of both LightGBM and CatBoost.

Bauder et. al [6] focused on the detection of Medicare Part B provider fraud which involves fraudulent activities, such as patient abuse or neglect and billing for services not rendered, perpetrated by providers and other entities who have been excluded from participating in Federal healthcare programs. They discuss Part B data processing and describe a unique process for mapping fraud labels with known fraudulent providers. The labeled big dataset is highly imbalanced with a very limited number of fraud instances. In order to combat this class imbalance, they generate seven class distributions and assess the behavior and fraud detection performance of six different machine learning methods. These results show that RF100 using a 90:10 class distribution is the best learner with a 0.87302 AUC. Moreover, learner behavior with the 50:50 balanced class distribution is similar to more imbalanced distributions which keep more of the original data. Based on the performance and significance testing results, they posit that retaining more of the majority class information leads to better Medicare Part B fraud detection performance over the balanced datasets across the majority of learners.

Herland et. al [7] focused on the detection of Medicare fraud using the following CMS datasets: (1) Medicare Provider Utilization and Payment Data: Physician and Other Supplier (Part B), (2) Medicare Provider Utilization and Payment Data: Part D Prescriber (Part D), and (3) Medicare Provider Utilization and Payment Data: Referring Durable Medical Equipment, Prosthetics, Orthotics and Supplies (DMEPOS). Additionally, they create a fourth dataset which is a combination of the three primary datasets. They discuss data processing for all four datasets and the mapping of real-world provider fraud labels using the List of Excluded Individuals and Entities (LEIE) from the Office of the Inspector General. This exploratory analysis on Medicare fraud detection involves building and assessing three learners on each dataset. Based on the Area under the Receiver Operating Characteristic (ROC) Curve performance metric, these results show that the Combined dataset with the Logistic Regression (LR) learner yielded the best overall score at 0.816, closely followed by the Part B dataset with LR at 0.805. Overall, the Combined and Part B datasets produced the best fraud detection performance with no statistical difference between these datasets, over all the learners. Therefore, based on these results and the assumption that there is no way to know within which part of Medicare a physician will commit fraud, they suggest using the Combined dataset for detecting fraudulent behavior when a physician has submitted payments through any or all Medicare parts evaluated in this study.

Arunkumar et. al [8] provides an extensive study of detecting fraudulent claims in healthcare insurance by leveraging machine learning algorithms. By using the publicly available medicare dataset, they are able to classify as fraud and non-fraud providers. Moreover, synthetically minority oversampling technique is used to avoid the class imbalance problem. Furthermore, a hybrid approach is used which is based on clustering and classification. Additionally, they have used other machine learning algorithms to check the efficiency of the best-suited algorithm.

Chen et. al [9] developed a framework of automatic medical fraud detection (AMFD) which can be deployed in healthcare industry. To address the issue that the medical fraud labels are insufficient in both size and classes for training a good AMFD model, this work proposes a novel Variational AutoEncoder-based Relational Model (VAERM) which can simultaneously exploit Patient-Doctor relational network and one-class fraud labels to improve the fraud detection. Then, the proposed VAERM coupled with active learning strategy can assist healthcare industry experts to conduct cost-effective fraud investigation. Finally, they propose an online model updating method to reduce the computation and memory requirement while preserving the predictive performance. The proposed framework is tested in a real-world dataset and it empirically outperforms the state-of-the-art methods in both automatic fraud detection and fraud investigation tasks.

Yao et. al [10] used the Bagging algorithm to build a Medicare fraud detection model. The Gradient Boost Tree, XGBoost, CatBoost, and DTC models, are proven effective in past studies, and are used as the base models to construct the Medicare fraud detection model. They proposed the Bagging algorithm based on the weighted threshold method named WTBagging and made ten model combinations using Bagging and WTBagging algorithms. The data are cleaned and sampled to construct three datasets with different class distributions. The 5-fold cross-validation process was applied to the model training and repeated ten times, and the F1 value was the performance metric to evaluate the model combination.

The results show that the model combinations of the WTBagging achieved the highest F1 values under all datasets.

Herland et. al [11] focused specifically on Medicare, utilizing three 'Big Data' Medicare claims datasets with real-world fraudulent physicians. They create a training and test dataset for all three Medicare parts, both separately and combined, to assess fraud detection performance. To emulate class rarity, which indicates particularly severe levels of class imbalance, they generate additional datasets, by removing fraud instances, to determine the effects of rarity on fraud detection performance. Before a machine learning model can be distributed for real-world use, a performance evaluation is necessary to determine the best configuration (e.g., learner, class sampling ratio) and whether the associated error rates are low, indicating good detection rates. With this research, they demonstrated the effects of severe class imbalance and rarity using a training and testing (Train Test) evaluation method via a hold-out set, and provide these recommendations based on the supervised machine learning results. Additionally, they repeat the same experiments using Cross-Validation, and determine it is a viable substitute for Medicare fraud detection. For machine learning with the severe class imbalance datasets, they founded that, as expected, fraud detection performance decreased as the fraudulent instances became rarer. They applying Random Under sampling to both Train Test and Cross-Validation, for all original and generated datasets, in order to assess potential improvements in fraud detection by reducing the adverse effects of class imbalance and rarity. Helmut Farbmacher et. al [12] develop a deep learning model that can handle these challenges by adapting methods from text classification. Using a large dataset from a private health insurer in Germany, they show that the model they propose outperforms a conventional machine learning model. With the rise of digitalization, unstructured data with characteristics similar to ours will become increasingly common in applied research, and methods to deal with such data will be needed.

3. Proposed System

The application of fraudulent Medicare provider detection technologies and strategies has wide-ranging benefits, including financial savings, enhanced patient safety, improved healthcare resource allocation, and the preservation of public trust in healthcare systems. As technology continues to advance, these applications are likely to become even more sophisticated and effective in combating healthcare fraud.

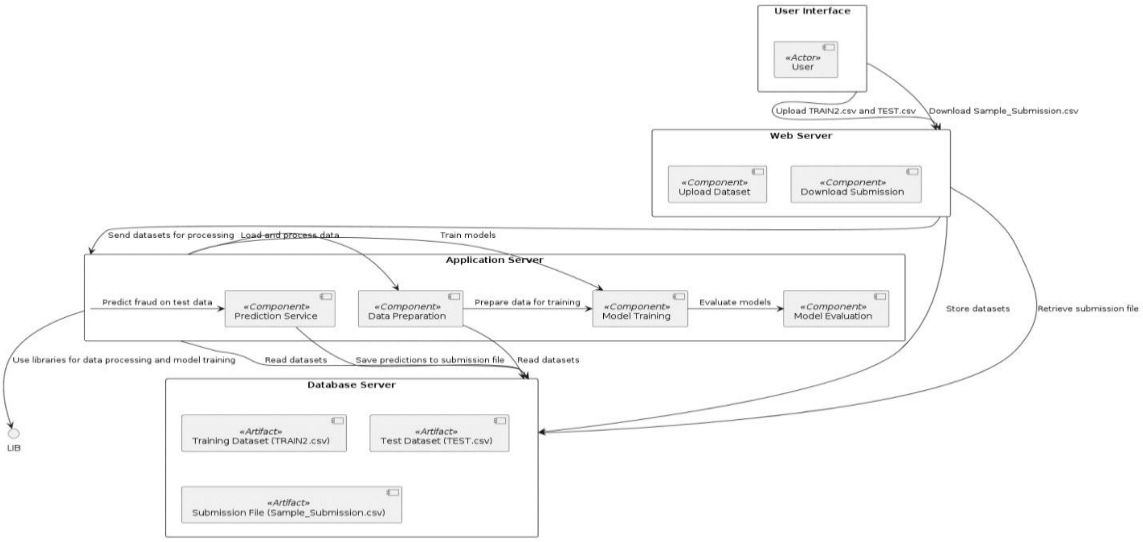


Figure 1: System architecture of proposed ML-driven detection of fraudulent medical providers.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 13.Iss 04, 2024

Figure 1 shows the proposed system model. The detailed operation of system model described as follows:

Step 1. Dataset: In healthcare fraud detection, you typically have a dataset containing information about healthcare providers. This dataset includes features like provider characteristics, billing patterns, services offered, and historical data.

Step 2. Data Preprocessing: Remove or impute missing values and handle outliers to ensure data quality. Choose relevant features and possibly create new ones to improve model performance. Convert categorical variables into numerical format (e.g., one-hot encoding or label encoding). Divide the data into training and testing sets for model evaluation.

Step 3. Apply Logistic Regression Model: A simple and interpretable algorithm for binary classification. It models the probability that a provider is fraudulent. Good for initial exploration of the problem.

Step 4: Apply Decision Tree Classifier: Builds a tree-like structure to make decisions based on feature values. Can capture non-linear relationships in the data. Prone to overfitting, but this can be mitigated with techniques like pruning.

3.1 DTC Algorithm

DTC is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model. As the name suggests, "DTC is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." Instead of relying on one decision tree, the DTC takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output. The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.

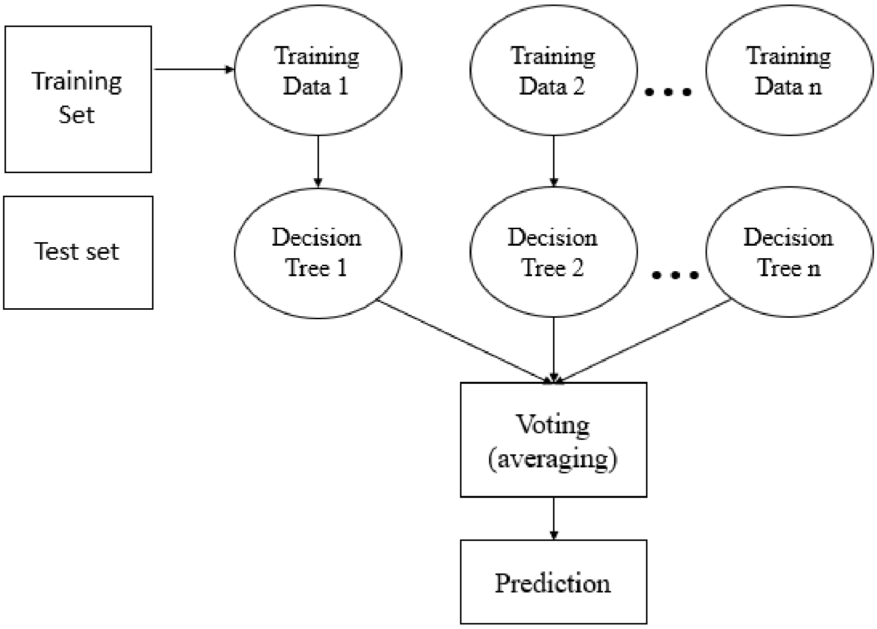


Figure 2: DTC algorithm.

DTC algorithm

Step 1: In DTC n number of random records are taken from the data set having k number of records.

Step 2: Individual decision trees are constructed for each sample.

Step 3: Each decision tree will generate an output.

Step 4: Final output is considered based on Majority Voting or Averaging for Classification and regression respectively.

Important Features of DTC

- **Diversity**- Not all attributes/variables/features are considered while making an individual tree, each tree is different.
- **Immune to the curse of dimensionality**- Since each tree does not consider all the features, the feature space is reduced.
- **Parallelization**-Each tree is created independently out of different data and attributes. This means that we can make full use of the CPU to build DTCs.
- **Train-Test split**- In a DTC we don't have to segregate the data for train and test as there will always be 30% of the data which is not seen by the decision tree.
- **Stability**- Stability arises because the result is based on majority voting/ averaging.

Assumptions for DTC

Since the DTC combines multiple trees to predict the class of the dataset, it is possible that some decision trees may predict the correct output, while others may not. But together, all the trees predict the correct output. Therefore, below are two assumptions for a better DTC classifier:

- There should be some actual values in the feature variable of the dataset so that the classifier can predict accurate results rather than a guessed result.
- The predictions from each tree must have very low correlations.

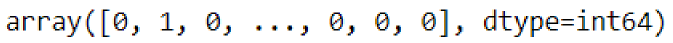
Below are some points that explain why we should use the DTC algorithm

- It takes less training time as compared to other algorithms.
- It predicts output with high accuracy, even for the large dataset it runs efficiently.
- It can also maintain accuracy when a large proportion of data is missing.

4. Results and discussion

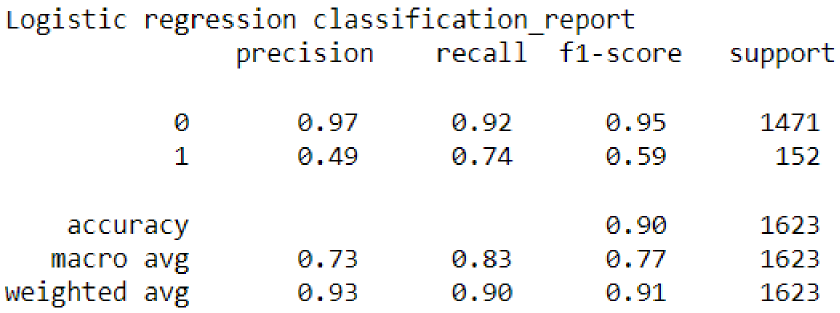
Figure 3 shows the representation of the array containing the target variables of the dataset. In the context of Medicare fraud detection, this array likely holds the labels indicating whether a provider is potentially fraudulent or not.

Figure 4 provides the detailed results of the classification report for the logistic regression model. The classification report includes important metrics such as precision, recall, F1-score, and support for each class. Here, it evaluates how well the logistic regression model is performing at identifying potential Medicare fraud.



```
array([0, 1, 0, ..., 0, 0, 0], dtype=int64)
```

Figure 3: Array of target variables of a dataset



```
Logistic regression classification_report
              precision    recall  f1-score   support

           0       0.97      0.92      0.95      1471
           1       0.49      0.74      0.59       152

    accuracy                           0.90      1623
   macro avg       0.73      0.83      0.77      1623
weighted avg       0.93      0.90      0.91      1623
```

Figure 4: classification report of Logistic regression

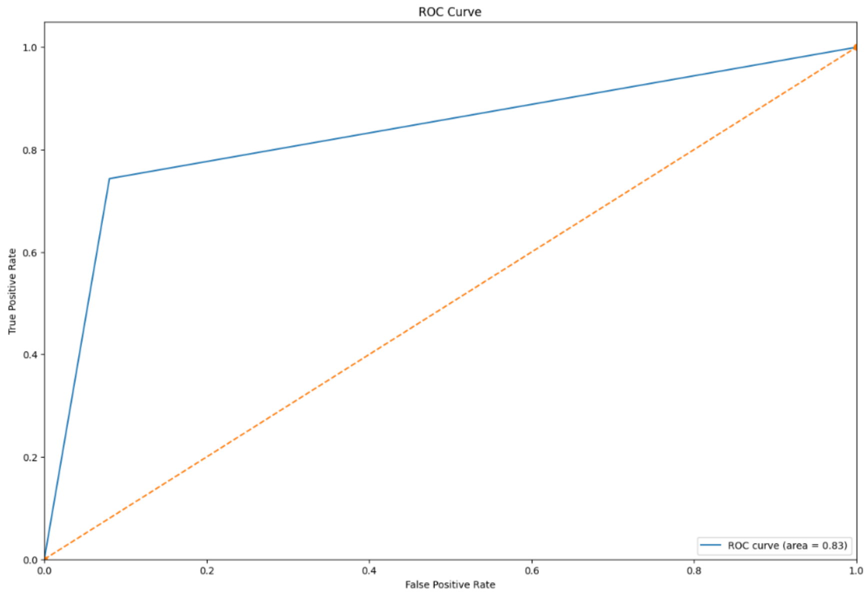
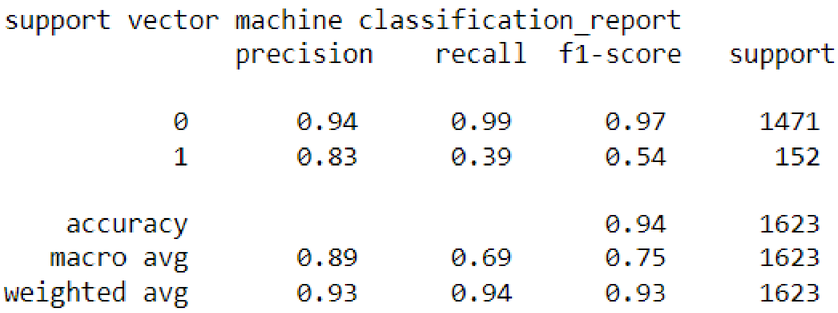


Figure 5: ROC curve for Logistic regression

Figure 5 displays the Receiver Operating Characteristic (ROC) curve for the logistic regression model. The ROC curve is a graphical representation of the true positive rate against the false positive rate. It's used to evaluate the performance of a binary classification model, and the area under the curve (AUC) can indicate how well the model is distinguishing between the two classes. Figure 6 provides the detailed results of the classification report for the support vector machine (SVM) model. Like Figure 4, it includes metrics such as precision, recall, F1-score, and support for each class. It evaluates how well the SVM model is performing at identifying potential Medicare fraud. Figure 7 displays the ROC curve for the support vector machine (SVM) model. Similar to Figure 5, it's a graphical representation of the true positive rate against the false positive rate. It's used to evaluate how well the SVM model is distinguishing between the two classes.



```
support vector machine classification_report
              precision    recall  f1-score   support

           0       0.94      0.99      0.97      1471
           1       0.83      0.39      0.54       152

    accuracy                           0.94      1623
   macro avg       0.89      0.69      0.75      1623
weighted avg       0.93      0.94      0.93      1623
```

Figure 6: classification report of support vector machine

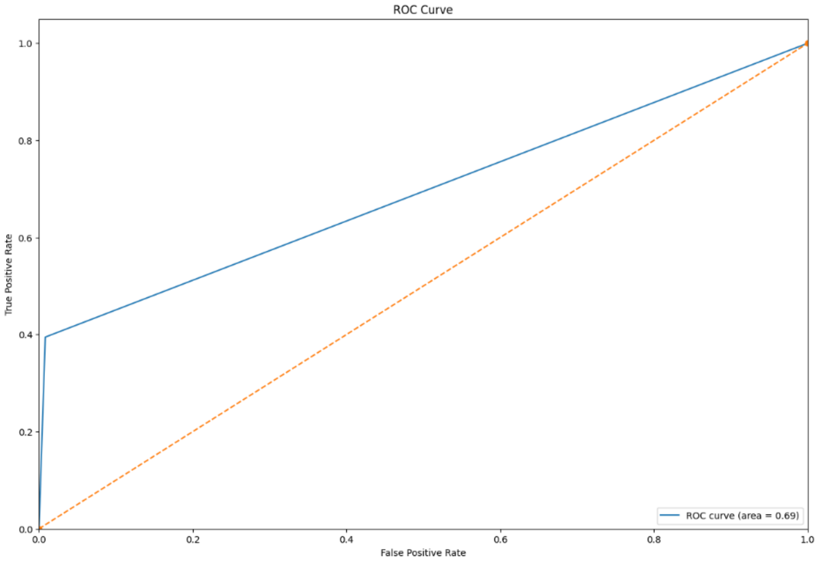


Figure 7: Roc curve for support vector machine algorithm

5. Conclusion

In conclusion, detecting fraudulent Medicare providers is a critical task in healthcare to ensure the integrity of the system and prevent financial losses. Start with a comprehensive dataset containing provider information, billing records, and historical data. Clean and preprocess the data, handling missing values, outliers, and encoding categorical variables. Split the data into training and testing sets. Consider using machine learning algorithms like Logistic Regression and Decision Trees for binary classification tasks. Choose the algorithm that best suits your dataset and problem requirements. Evaluate model performance using metrics such as accuracy, precision, recall, and F1-score. The confusion matrix provides a detailed breakdown of model predictions. The DTC classifier resulted in superior performance over existing models. Experiment with different algorithms and hyperparameters to improve model performance. Employ techniques like cross-validation for robust performance estimation. Continuously monitor the model's performance and adapt to changing fraud patterns. Be prepared to retrain the model with new data to stay effective over time.

References

[1] Lakshman Narayana Vejendla and A Peda Gopi, (2019),” Avoiding Interoperability and Delay in Healthcare Monitoring System Using Block Chain Technology”, Revue d'Intelligence Artificielle, Vol. 33, No. 1, 2019, pp.45-48.

[2] Gopi, A.P., Jyothi, R.N.S., Narayana, V.L. et al. (2020), “Classification of tweets data based on polarity using improved RBF kernel of www.jespublication.com PageNo:482 SVM”. Int. j. inf. tecnol. (2020)

[3] Lakshman Narayana Vejendla and A Peda Gopi, (2017),” Visual cryptography for gray scale images with enhanced security mechanisms”, Traitement du Signal, Vol.35, No.3-4, pp.197-208. DOI: 10.3166/ts.34.197-208

[4] Herland, M., Bauder, R.A. & Khoshgoftaar, T.M. Approaches for identifying U.S. medicare fraud in provider claims data. Health Care Manag Sci 23, 2–19 (2020). https://doi.org/10.1007/s10729-018-9460-8

[5] Hancock, J.T., Khoshgoftaar, T.M. Hyperparameter Tuning for Medicare Fraud Detection in Big Data. SN COMPUT. SCI. 3, 440 (2022). https://doi.org/10.1007/s42979-022-01348-x

[6] Bauder, R.A., Khoshgoftaar, T.M. The effects of varying class distribution on learner behavior for medicare fraud detection with imbalanced big data. Health Inf Sci Syst 6, 9 (2018). https://doi.org/10.1007/s13755-018-0051-3

[7] Herland, M., Khoshgoftaar, T.M. & Bauder, R.A. Big Data fraud detection using multiple medicare data sources. J Big Data 5, 29 (2018). https://doi.org/10.1186/s40537-018-0138-3

[8] Arunkumar, C., Kalyan, S., Ravishankar, H. (2021). Fraudulent Detection in Healthcare Insurance. In: Sengodan, T., Murugappan, M., Misra, S. (eds) Advances in Electrical and Computer Technologies. ICAECT 2020. Lecture Notes in Electrical Engineering, vol 711. Springer, Singapore. https://doi.org/10.1007/978-981-15-9019-1_1

[9] J. Chen, X. Hu, D. Yi, J. Li and M. Alazab, "A Variational AutoEncoder-Based Relational Model for Cost-Effective Automatic Medical Fraud Detection," in IEEE Transactions on Dependable and Secure Computing, 2022, doi: 10.1109/TDSC.2022.3187973.

[10] J. Yao, S. Yu, C. Wang, T. Ke and H. Zheng, "Medicare Fraud Detection Using WTBagging Algorithm," 2021 7th International Conference on Computer and Communications (ICCC), 2021, pp. 1515-1519, doi: 10.1109/ICCC54389.2021.9674545.

[11] Herland, M., Bauder, R.A. & Khoshgoftaar, T.M. The effects of class rarity on the evaluation of supervised healthcare fraud detection models. J Big Data 6, 21 (2019). https://doi.org/10.1186/s40537-019-0181-8

[12] Helmut Farbmacher, Leander Löw, Martin Spindler, An explainable attention network for fraud detection in claims management, Journal of Econometrics, Volume 228, Issue 2, 2022, Pages 244-258, ISSN 0304-4076, https://doi.org/10.1016/j.jeconom.2020.05.021.