Health Insurance Fraud Detection using Machine Learning Techniques

Subject: Intelligent Database Systems

Subject code: BCD3006

Slot: A2+TA2

Faculty: Swetha NG

Group Members:

Syed Qasim Kaif (20BDS0343)

Riyaz Mohammad Arbaz (20BDS0274)

Sai Charan A (20BDS0354)

Syed Sha Suheb (20BDS0394)

Table of Content

Abstract:

Inappropriate payments by insurance organizations or third-party payers occur because of errors, abuse, and fraud. It is estimated that approximately 10% of medical expenditures are wasted in medical fraud and abuse. The scale of this problem is large enough to make it a priority issue for health systems. Traditional methods of detecting health care fraud and abuse are time-consuming and inefficient. Combining automated methods and statistical knowledge leads to the emergence of a new interdisciplinary branch of science that is named Knowledge Discovery from Databases (KDD). Data mining and machine learning is a core of the KDD process. Despite putting up various technologies and strategies to fight fraud such as planned, targeted, audits and random audits, whistle blowing, and biometric systems, fraud in claims have continued to be a challenge in most of the health insurance providers across the world. Fraud detection in health insurance companies, is much needed in developed and undeveloped countries to reduce loss of money and resources and in return improve the service delivery to patients.

- This project tried to analyze the appropriateness of data mining techniques in detecting fraudulent health insurance claims.
- The goal of this project is to " predict the potentially fraudulent providers " based on the claims filed by them.
- Along with this, we will also discover important variables helpful in detecting the behavior of potentially fraud providers.
- further, we will study fraudulent patterns in the provider's claims to understand the future behavior of providers. To achieve our goal classification models were used to guide the entire knowledge discovery process.
- Classification and regression algorithms such as Random Forest, SVM, Naïve Bayes, Decision Tree, Gradient booster etc. were used to build predictive models.
- Several Experiments were conducted, and the resulting models shows that SVM works well among the other algorithms in predicting fraud claims with an accuracy of 91.7%.

Introduction

Need for Problem Statement

Health insurance fraud is a significant problem that affects the healthcare industry and the wider economy. It refers to any intentional misrepresentation or deception by an individual or organization to obtain payment or benefit from an insurance company or a government healthcare program. Health insurance fraud can take various forms, including billing for services not provided, overbilling for services, and providing unnecessary or substandard services. The detection of health insurance fraud is a complex process that requires sophisticated techniques and tools. The healthcare industry involves numerous stakeholders, including healthcare providers, patients, insurance companies, and government agencies. Each of these stakeholders has different motivations and interests, making it challenging to detect fraud.

- One of the primary challenges in detecting health insurance fraud is the sheer scale and complexity of the healthcare industry.
- Another critical challenge in detecting health insurance fraud is the use of sophisticated techniques and tools.
- The third challenge in detecting health insurance fraud is the legal and ethical issues involved.
- The use of sensitive healthcare data for fraud detection must be done in a way that protects patient privacy and confidentiality. A problem statement helps to identify the legal and ethical issues involved and develop strategies that comply with privacy laws and regulations.
- Another significant challenge in detecting health insurance fraud is the need to balance fraud detection with the efficient payment of legitimate claims. A problem statement helps to develop strategies that balance the detection of fraud with the efficient payment of legitimate claims.
- Finally, the need for a problem statement in health insurance fraud is critical to develop effective prevention strategies. A problem statement helps to identify the root causes of health insurance fraud and develop strategies to address them proactively.

Basic terminologies

Health insurance fraud occurs when someone intentionally makes false statements or deceives an insurer to receive payments or benefits that they are not entitled to. Here are some basic terminologies related to health insurance fraud:

1. Fraudulent Claims

   Claims that are submitted by healthcare providers, patients, or others that contain false/misleading information with the intent to receive payment for services or treatments they provide.

2. Kickbacks

   Kickbacks are payments or gifts given to individuals or entities as an incentive to refer patients or business to a particular healthcare provider or facility.

3. Phantom billings

   It is a type of health insurance fraud in which healthcare providers bill for services or treatments that were never provided.

4. Upcoding

   It occurs when healthcare providers use billing codes to bill for more expensive procedures or services than were provided.

5. Unbundling

   This is a type of fraud in which healthcare providers bill separately for services that are typically billed together as part of a single procedure.

6. Double Billing

   Occurs when healthcare providers bill more than once for the same service or treatment.

7. Falsifying Medical Records

   Falsifying medical records involves altering or fabricating medical records to support false claims or to cover up fraudulent activities.

8. Identity Theft

   Identity theft occurs when someone steals another person's personal information and uses it to obtain medical services or prescriptions under the victim's name. This is a form of healthcare fraud that can result in financial harm to the victim.

Outcomes

In conclusion, health insurance fraud is a complex issue that can involve many different types of fraudulent activities. By understanding these basic terminologies, individuals can become more aware of the types of fraud that exist and take steps to protect themselves from becoming victims of healthcare fraud. It is essential to report any suspicious activity to the appropriate authorities to help combat this widespread problem.

In conclusion, health insurance fraud is a significant problem that requires sophisticated tools and techniques to detect and prevent. The need for a problem statement in health insurance fraud is critical to identify the challenges involved and develop effective strategies to detect and prevent fraud. By addressing these challenges, insurance companies and government agencies can protect their customers, maintain their reputation, and reduce the financial losses associated with health insurance fraud.

Literature Survey

The research paper identifies that imbalanced data is a common issue in healthcare insurance fraud detection, with the minority class being significantly smaller in number [1]. Research and validation are needed to ensure the accuracy, reliability, and safety of AI-based applications in the COVID-19 medical [2]. Proposes a multidimensional data model and analysis techniques for predicting healthcare fraud in Medicaid [3]. A systematic study of statistical methods for healthcare fraud detection [6]. The article gives a thorough assessment of the manifestations and contributing causes of health insurance fraud over a fourteen-year period [7]. Rapid data processing using clustering algorithms enables real-time fraud detection. Clustering techniques may produce false positives or false negatives (valid claims being labelled as fraudulent) (fraudulent claims being missed) [9]. Increased precision: By seeing trends and anomalies in data on health insurance claims, AI and blockchain technology can assist detect fraud with greater precision. The accuracy of clustering techniques depends significantly on the data's quality. Results may be erroneous if the data is noisy or lacking [10]. Collaboration may be strengthened because of the study to combat health insurance fraud in China. Insurance firms, healthcare providers, and regulators can all work together more effectively [11]. The manual vetting of claims is time-consuming and can lead to delays in service delivery, which affects both the insurance companies and the patients. The proposed DSS can help to reduce the computational time of claims processing while increasing classification accuracy [12]. The evolution of data types used in fraud detection practices has gone from basic quantitative data to multi-source unstructured data, and the trend is to use more panoramic data to comprehensively detect fraud activities [13]. A medical insurance dataset from the KAGGLE repository was used to train and test various regression models, including Linear Regression, Ridge Regressor, Support Vector Regression, XGBoost, Stochastic Gradient Boosting, Decision Tree, Random Forest Regressor, k-Nearest Neighbors, and Multiple Linear Regression [14]. The model uses a focal-loss function to adapt to data imbalance and a relative probability score to measure performance [15].

Title & Author	Relevant Findings	limitations
Imbalanced classification problems: Systematic study and challenges in healthcare insurance fraud detection. Mary, A. J., & Claret, S. A.	Performance evaluation metrics such as precision, recall, F1-score, AUPRC, and AUROC are recommended. Sampling techniques and ensemble methods are proposed as potential solutions	lack of standardized benchmark datasets for healthcare insurance fraud detection, making it challenging to compare results across studies
Application of artificial intelligence in COVID-19 medical area: a systematic review. Chang, Z., Zhan, Z., Zhao, Z., You, Z., Liu, Y., Yan, Z., ... & Zhao, L.	findings on the application of artificial intelligence in the medical area of COVID-19 through a systematic review. The paper identifies the potential of AI in various areas such as diagnosis, prognosis, treatment, and monitoring of COVID-19.	lack of standardized benchmark datasets, limited studies with small sample sizes, potential bias in data collection, and the need for validation in real-world clinical settings.
Predicting healthcare fraud in Medicaid: a multi-dimensional data model and analysis techniques for fraud detection. Thornton, D., Mueller, R. M., Schoutsen, P., & Van Hillegersberg, J.	This highlights the importance of incorporating multiple dimensions of data, such as medical, financial, and demographic information, for accurate fraud detection	Lack of real-world data validation and the need for further investigation on the scalability and generalizability of the proposed approach in different healthcare settings.
A novel fraud detection and prevention method for healthcare claim processing using machine learning and blockchain technology. Amponsah, A. A., Adekoya, A. F., & Weyori, B. A.	A novel fraud detection and prevention method for healthcare claim processing using machine learning and blockchain technology.	lack of standardized benchmark datasets and potential challenges in implementing blockchain technology in real-world healthcare settings.
Using data mining to detect health care fraud and abuse: a review of literature. Global journal of health science. Joudaki, H., Rashidian, A., Minaei-Bidgoli, B., Mahmoodi, M., Geraili, B., Nasiri, M., & Arab, M.	Use of data mining techniques for healthcare fraud and abuse detection. Data mining approaches, such as decision trees, support vector machines, and ensemble methods, can be effective in detecting fraud and abuse.	Lack of standardized benchmark datasets, challenges in handling imbalanced data, and the need for domain-specific considerations.

A survey on statistical methods for health care fraud detection. Health care management. Li, J., Huang, K. Y., Jin, J., & Shi, J.	Findings include the use of statistical techniques such as clustering, anomaly detection, and regression for fraud detection.	Lack of standardized benchmark datasets, challenges in feature selection, and potential false positives/negatives. The paper emphasizes the need for domain-specific considerations.
Fourteen years of manifestations and factors of health insurance fraud, 2006–2020 Villegas-Ortega, J., Bellido-Boza, L., & Mauricio, D.	The research offers a thorough overview of the different fraud schemes that affect the health insurance industry, such as prescription fraud, billing fraud, and insurance identity theft.	Most of the research on health insurance fraud, according to the authors, is done in the United States, which restricts the applicability of the findings to other nations and healthcare systems.
Health care insurance fraud detection using blockchain. Saldamli, G., Reddy, V., Bojja, K. S., Gururaja, M. K., Doddaveerappa, Y., & Tawalbeh, L.	Highlights that blockchain technology can enhance fraud detection by providing a transparent and immutable record of transactions. stakeholders, improving fraud detection accuracy.	Need for standardization of blockchain implementation, potential scalability issues, legal and regulatory challenges.
Decision Support System (DSS) for Fraud Detection in Health Insurance Claims Using Genetic Support Vector Machines (GSVMs) Robert A. Sowah**,** Marcellinus Kuuboore, Abdul Ofoli, Samuel Kwofie,Louis Asiedu, Koudjo M. Koumadi, and Kwaku O. Apeadu	The study used a real-world dataset from the National Health Insurance Scheme in Ghana to evaluate the performance of the proposed method. The results showed that the GSVM model with the Radial Basis Function (RBF) kernel had the best performance with an average accuracy rate of 87.91%.	the proposed model was evaluated using only one dataset. Another limitation is the skewed nature of the claimed dataset towards certain medical specialties, which may affect the performance.
Intelligent financial fraud detection practices in post-pandemic era Xiaoqian Zhu, Xiang Ao, Zidi qin, Yanpeng chang, Yang Liu, Qing He, Jianping Li	Deep learning (DL) systems, particularly graph-based detection approaches such as Graph Neural Networks (GNN), are becoming popular in fraud detection due to their versatility and ability to analyze multi-source data.	challenges and potential directions for future development, particularly in achieving a comprehensive understanding and accurate identification of fraud activities and addressing the limitations of current models in the context of financial fraud detection.

Proposed models

We have used both supervised and unsupervised approaches for detecting fraud and compared and evaluated the various models and algorithms.

Here is the list of all algorithms that we have used in our project.

Random forest

Random Forest is a 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.

Working:

Random Forest works in two-phase first is to create the random forest by combining N decision tree, and second is to make predictions for each tree created in the first phase.

Step-1: Select random K data points from the training set.

Step-2: Build the decision trees associated with the selected data points (Subsets).

Step-3: Choose the number N for decision trees that you want to build.

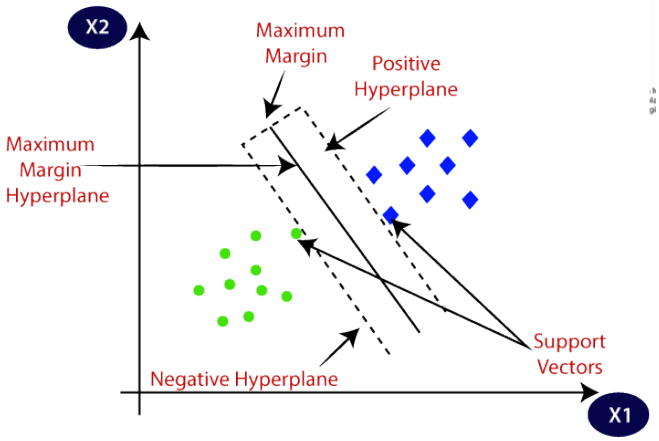
Step-4: Repeat Step 1 & 2.

Step-5: For new data points, find the predictions of each decision tree, and assign the new data points to the category that wins the majority votes.

SVM Algorithm

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning.

The goal of the SVM algorithm is to create the best line or decision boundary that can



segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane.

SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called as support vectors, and hence algorithm is termed as Support Vector Machine.

Naive Bayes

Naïve Bayes algorithm is a supervised learning algorithm, which is based on Bayes theorem and used for solving classification problems.

- It is mainly used in text classification that includes a high-dimensional training dataset.
- Naïve Bayes Classifier which helps in building the fast machine learning models that can make quick predictions.
- It is a probabilistic classifier, which means it predicts based on the probability of an object.
- Some popular examples of Naïve Bayes Algorithm are spam filtration, Sentimental analysis, and classifying articles.

Working:

1. Convert the given dataset into frequency tables.
2. Generate Likelihood table by finding the probabilities of given features.
3. Now, use Bayes theorem to calculate the posterior probability.

Applications

- It is used for Credit Scoring, also used in medical data classification.
- It can be used in real-time predictions because Naïve Bayes Classifier is an eager learner. Also used in Text classification such as Spam filtering and Sentiment analysis.

Decision Tree

It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome.

In a Decision tree, there are two nodes, which are the Decision Node and Leaf Node. Decision nodes are used to make any decision and have multiple branches, whereas Leaf nodes are the output of those decisions and do not contain any further branches.

The decisions or the test are performed based on features of the given dataset.

It is a graphical representation for getting all the possible solutions to a problem/decision based on given conditions.

It is called a decision tree because, like a tree, it starts with the root node, which expands on further branches and constructs a tree-like structure.To build a tree, we use the CART algorithm, which stands for Classification and Regression Tree algorithm.

A decision tree simply asks a question and based on the answer (Yes/No), it further split the tree into subtrees.

Decision Tree Terminologies

Root Node: Root node is from where the decision tree starts. It represents the entire dataset, which further gets divided into two or more homogeneous sets.

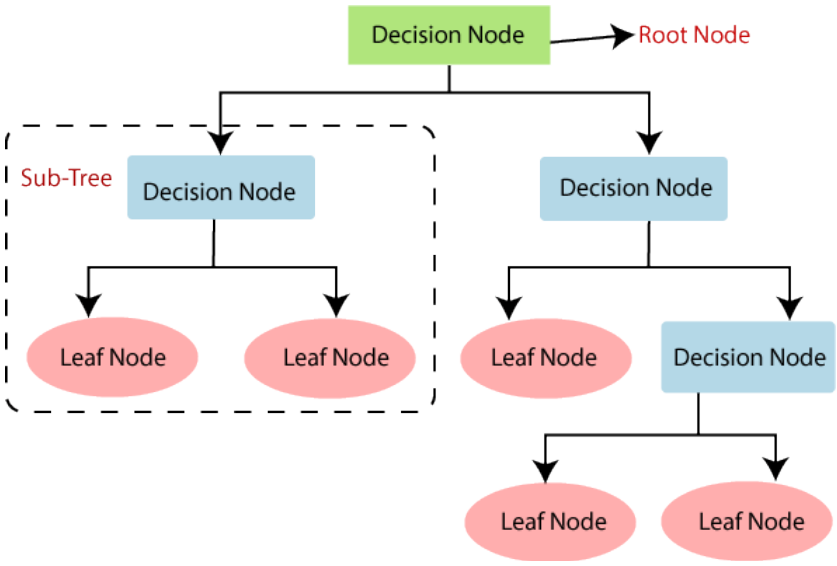
Leaf Node: Leaf nodes are the final output node, and the tree cannot be segregated further after getting a leaf node.

Splitting: Splitting is the process of dividing the decision node/root node into sub-nodes according to the given conditions.

Branch/Sub Tree: A tree formed by splitting the tree.

Pruning: Pruning is the process of removing the unwanted branches from the tree.

Parent/Child node: The root node of the tree is called the parent node, and other nodes are called the child nodes.



Working

Step-1: Begin the tree with the root node, says S, which contains the complete dataset.

Step-2: Find the best attribute in the dataset using Attribute Selection Measure (ASM).

Step-3: Divide the S into subsets that contains possible values for the best attributes.

Step-4: Generate the decision tree node, which contains the best attribute.

Step-5: Recursively make new decision trees using the subsets of the dataset created in step -3. Continue this process until a stage is reached where you cannot further classify the nodes and called the final node as a leaf node.

Reasons of using decision Tree

There are various algorithms in Machine learning, so choosing the best algorithm for the given dataset and problem is the main point to remember while creating a machine learning model. Below are the two reasons for using the Decision tree:

- Decision Trees usually mimic human thinking ability while deciding, so it is easy to understand.
- The logic behind the decision tree can be easily understood because it shows a tree-like structure.

Gradient Boosting Algorithm:

Gradient boosting is a method standing out for its prediction speed and accuracy, particularly with large and complex datasets. From Kaggle competitions to machine learning solutions for business, this algorithm has produced the best results. We already know that errors play a major role in any machine learning algorithm. There are mainly two types of error, bias error, and variance error. Gradient boost algorithm *helps us minimize bias error* of the model.

the main idea behind this algorithm is to build models sequentially and these subsequent models try to reduce the errors of the previous model. But how do we do that? How do we reduce the error? This is done by building a new model on the errors or residuals of the previous model.

When the target column is continuous, we use Gradient Boosting Regressor whereas when it is a classification problem, we use Gradient Boosting Classifier. The only difference between the two is the *"**Loss function**"*. The objective here is to minimize this loss function by adding weak learners using gradient descent. Since it is based on loss function hence for regression problems, we'll have different loss functions like Mean squared error (MSE) and for classification, we will have different for e.g., log-likelihood.

Results And Discussions

The total of 6 data sets were taken representing different features.

Introduction to the Dataset

For the purpose of this project, we are considering Inpatient claims, Outpatient claims and Beneficiary details of each provider. Lets s see their details :

A) Inpatient Data

This data provides insights about the claims filed for those patients who are admitted in the hospitals. It also provides additional details like their admission and discharge dates and admit d diagnosis code.

B) Outpatient Data

This data provides details about the claims filed for those patients who visit hospitals and not admitted in it.

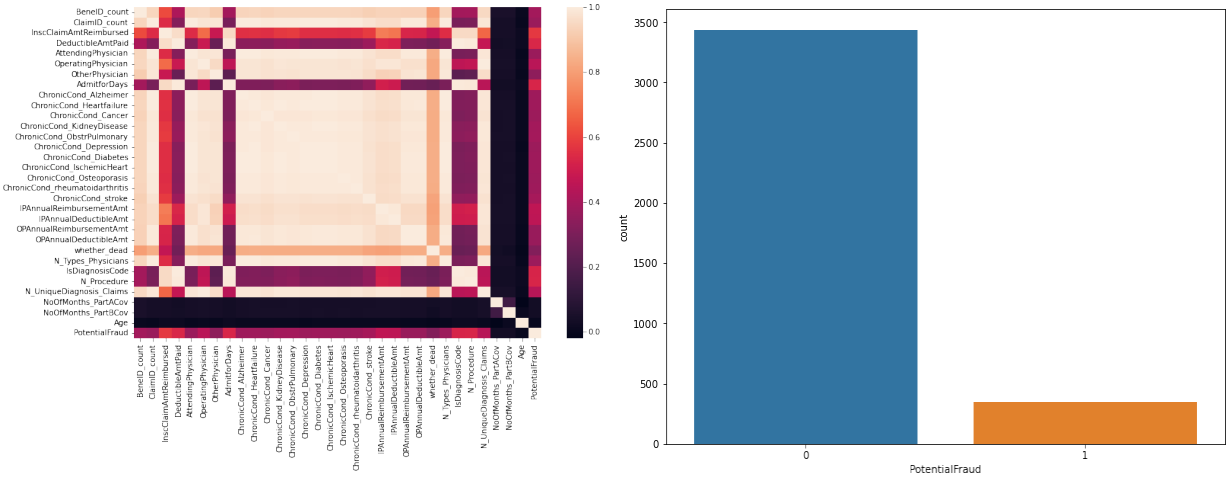
C) Beneficiary Details Data

This data contains beneficiary KYC details like health conditions, region they belong to etc. the initial data set was provided with BeneID, DOB etc. These features and then preprocessing was performed on beneficiary, inpatient and outpatient data sets followed by merging of the datasets into one single and final data set on which modelling was done.

Data preparation

In this phase the appropriate datasets that were best fitted to perform modelling were chosen. The data was then cleaned, integrated, and formatted in order for it to be fit for use. Data cleaning activities included removing duplicate records, correcting noisy data, filling the missing values by using estimation, removing the attributes and records that are irrelevant and not used for the data mining problem. The cleaned data was then further processed for dimensionality and numerosity reduction all this was done using the Python programming language on Jupyter Notebook Platform.

Finally, the final input data set for running the classification algorithms as training and testing data was generated.



From the above graphs we have a correlation matrix, we found that our data set is imbalanced as the number of 1's and zeroes are not proportionate hence we use sampling techniques to correct the imbalance data set.

Modeling

In this phase several modeling techniques were picked and applied the five basic techniques that we used were random forest, SVM, Decision tree, Naïve Bayes classifier and Graded Boosting Classifier.

Evaluation

The models that were prepared in the modeling phase were then evaluated and the best one was selected among them. The algorithms were tested on test data set to see how many of the test are classified as true positives and false positives. Then the performance of algorithm was evaluated taking into consideration the classification matrix, Accuracy, precision, Recall and F1 score.

Achieved Metrics:

Fraud is not detected: 0 (support = 1465)

Sampling methods	Accuracy	Precision	recall	F1-score	Accuracy
SVM-SMOTE	0.89716748	0.98	0.91	0.94	0.90
RANDOM FOREST-SMOTE	0.91194581	0.96	0.94	0.95	0.91
GRADIENT CLASSIFIER-SMOTE	0.91071428	0.97	0.92	0.95	0.91
NAÏVE BAYES-SMOTE	0.91379310	0.95	0.96	0.95	0.91
DECISION TREE-SMOTE	0.88793103	0.95	0.92	0.94	0.89
DECISION TREE-UNDER SAMPLING	0.43165024	0.97	0.38	0.55	0.43
SVM	0.775862068	0.98	0.77	0.86	0.78
RANDOM FOREST-OVER SAMPLING	0.898399014	0.98	0.91	0.94	0.90

Fraud is detected: 1 (support = 159)

Sampling methods	Accuracy	Precision	recall	F1-score	Accuracy
SVM-SMOTE	0.89716748	0.48	0.80	0.60	0.90
RANDOM FOREST-SMOTE	0.91194581	0.54	0.68	0.60	0.91
GRADIENT CLASSIFIER-SMOTE	0.91071428	0.53	0.78	0.63	0.91
NAÏVE BAYES-SMOTE	0.91379310	0.57	0.50	0.53	0.91
DECISION TREE-SMOTE	0.88793103	0.44	0.57	0.50	0.89
DECISION TREE-UNDER SAMPLING	0.43165024	0.14	0.89	0.24	0.43
SVM	0.775862068	0.28	0.83	0.42	0.78
RANDOM FOREST-OVER SAMPLING	0.898399014	0.49	0.79	0.60	0.90

Summary and conclusion

1. the project established that the main fraudulent activities perpetrated in health insurance claims are billing for more overpriced facilities or services than those that were truly administered, carrying out medically services that are not required, false claims, and forging a patient's diagnosis to explain some tests.

2. In this project the main classification techniques such as random forest,SVM, Naïve bayes etc were used. These techniques were chosen because they have been successfully applied in predictive analysis over time and they need a comparatively little effort from the users in preparing the data to solve scale contrasts between parameters. From the result analysis of our tested models the SVM classifier technique emerged as the best model in comparisons with other data mining techniques in terms of accuracy and other confusion parameters.

Conclusion

Fraud detection is a field that cannot rest, Fraudsters do exist and always will try new ways to perform frauds. Data mining discovers patterns not quite visible in data to convey some Knowledge. Fraud detection in health insurance companies in both developed and underdeveloped countries is in much needed as inability to discover fraudulent claims put them at risk of losing a lot of money and in turn affecting the service delivery to patients.

Concluding that the accuracy generated by naive bayes is much higher than all other models. But we choose SVM in general due to the future independence of feature.

Developing a mechanism to predict fraud is considered to be an achievement in health organizations and more such ideas should be supported and implemented in order to reduce fraud on a large scale.

Future works

Integration with fully functioning web application,Using transfer model on the trained model

References

1. Mary, A. J., & Claret, S. A. (2021, June). Imbalanced classification problems: Systematic study and challenges in healthcare insurance fraud detection. In 2021 5th International Conference on Trends in Electronics and Informatics (ICOEI) (pp. 1049-1055). IEEE.

2. Chang, Z., Zhan, Z., Zhao, Z., You, Z., Liu, Y., Yan, Z., ... & Zhao, L. (2021). Application of artificial intelligence in COVID-19 medical area: a systematic review. Journal of Thoracic Disease, 13(12), 7034.

3. Thornton, D., Mueller, R. M., Schoutsen, P., & Van Hillegersberg, J. (2013). Predicting healthcare fraud in medicaid: a multidimensional data model and analysis techniques for fraud detection. Procedia technology, 9, 1252-1264.

4. Amponsah, A. A., Adekoya, A. F., & Weyori, B. A. (2022). A novel fraud detection and prevention method for healthcare claim processing using machine learning and blockchain technology. Decision Analytics Journal, 4, 100122.

5. Joudaki, H., Rashidian, A., Minaei-Bidgoli, B., Mahmoodi, M., Geraili, B., Nasiri, M., & Arab, M. (2015). Using data mining to detect health care fraud and abuse: a review of literature. Global journal of health science, 7(1), 194.

6. Li, J., Huang, K. Y., Jin, J., & Shi, J. (2008). A survey on statistical methods for health care fraud detection. Health care management science, 11, 275-287.

7. Villegas-Ortega, J., Bellido-Boza, L., & Mauricio, D. (2021). Fourteen years of manifestations and factors of health insurance fraud, 2006–2020: a scoping review. Health & justice, 9, 1-23.

8. Saldamli, G., Reddy, V., Bojja, K. S., Gururaja, M. K., Doddaveerappa, Y., & Tawalbeh, L. (2020, April). Health care insurance fraud detection using blockchain. In 2020 Seventh international conference on software defined systems (SDS) (pp. 145-152). IEEE.

9.Peng, Y., Kou, G., Sabatka, A., Chen, Z., Khazanchi, D., & Shi, Y. (2006, October). Application of clustering methods to health insurance fraud detection. In 2006 International Conference on Service Systems and Service Management (Vol. 1, pp. 116-120). IEEE.

10. Kapadiya, K., Patel, U., Gupta, R., Alshehri, M. D., Tanwar, S., Sharma, G., & Bokoro, P. N. (2022). Blockchain and AI-Empowered Healthcare Insurance Fraud Detection: An Analysis, Architecture, and Future Prospects. IEEE Access, 10, 79606-79627.

11. Li, J., Lan, Q., Zhu, E., Xu, Y., & Zhu, D. (2022). A study of health insurance fraud in China and recommendations for fraud detection and prevention. Journal of Organizational and End User Computing (JOEUC), 34(4), 1-19.

iaoqian Zhu, Xiang Ao, Zidi Qin, Yanpeng Chang, Yang Liu, Qing He, Jianping Li,

Intelligent financial fraud detection practices in post-pandemic era,The Innovation,Volume 2, Issue 4,2021,100176,ISSN 2666-6758.

13. Zhang C, Xiao X, Wu C. Medical Fraud and Abuse Detection System Based on Machine Learning. International Journal of Environmental Research and Public Health. 2020; 17(19):7265.

14. Sowah, R. A., Kuuboore, M., Ofoli, A., Kwofie, S., Asiedu, L., Koumadi, K. M., & Apeadu, K. O. (2019). Decision support system (DSS) for fraud detection in health insurance claims using genetic support vector machines (GSVMs). Journal of Engineering, 2019.

15. ul Hassan, C. A., Iqbal, J., Hussain, S., AlSalman, H., Mosleh, M. A., & Sajid Ullah, S. (2021). A computational intelligence approach for predicting medical insurance cost. Mathematical Problems in Engineering, 2021, 1-13.