ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN2454-9940

www.ijasem.org

Vol 18, Issue 2, 2024

FRAUD DETECTION AND ANALYSIS FOR INSURANCE CLAIM USING MACHINE LEARNING

[1]MR.A N L KUMAR, [2]BOMMAREDDY TEJA REDDY

[1](Assistant Professor), MCA, Swarnandra College

[2]MCA, scholar, Swarnandra College

ABSTRACT

The therapeutic protection mechanism is particularly vulnerable to the pervasive and expensive fraud. Covering up or distorting facts in order to get social insurance benefits is one kind of unscrupulous protection. The protection guarantor or the protected could submit any number of different types of cheats. When it comes to health care, extortion is all down to dishonest social insurance providers. Associate in nursing research on extortion recognized evidence and exploitative instances is the dedication of this case misleading discovery. Consequently, information processing processes are used to detect the deception. Most of the time, essential based anomalies are used in conjunction with k-means, rule-based mining, affiliation rule bolstered appropriation calculations, and applied math call rules. These irregularities reveal the extortion in certified data. That being said, a lot may be accomplished by making use of different information processing processes. Trial findings from our technique are effective in human services deception, and the planned methodology has been evaluated based on protection information. This security data may also be subjected to other self-serving deception location methods.

1.INTRODUCTION

In recent years, data analytics has become more important to almost every sector of economic growth. There are tremendous prospects for data analysts in the healthcare industry, which accounts for a significant portion of the US economy&#39;s revenue. This data includes medical records, medications, insurance claims, provider details, and patient information, among other things. The annual cost of healthcare in the US exceeds three trillion dollars, and health insurance companies handle billions of claims. A typical healthcare reconciliation procedure involving many organizations is shown in Figure 1, which provides a clear flow diagram. Prior to administering any treatment, the provider&#39;s office verifies that the patient has sufficient financial resources, either from insurance or other sources. After then, the supplier of the service uses the results of the first exams to make a diagnosis. After then, the doctor or other medical professional will do testing on the patient, which may include more diagnostics or perhaps surgery. Typically, the patient&#39;s report will include these diagnoses and procedures as well as additional data such as demographics, personal information, and information about previous and current visits. Here is where most patients pay their copayments as outlined in their insurance policies and then check out. The next step is for a medical coder to review the patient&#39;s report, extract relevant data, and then generate a &quot;superbill&quot; that includes the provider&#39;s information, It is not surprising to see insurance companies dealing with false and fraudulent claims, considering the size of the healthcare business. Typical examples of healthcare fraud perpetrated by unscrupulous professionals include the following.

• Using erroneous diagnosis to justify unnecessary medical treatments.

• &quot;Upcoding,&quot; the practice of billing for expensive services or processes rather than the actual treatments themselves.

• Making assertions for processes that have not been carried out.

• Claiming insurance money for medically unwarranted treatments.

• A practice known as &quot;unbundling&quot; involves billing for individual steps in a process rather than as a whole.

• Claiming cosmetic operations or other non-covered treatments as medically essential in order to get insurance to pay for them.

Furthermore, when compared to other industries, healthcare is the most reluctant to share data. Additionally, it is not possible to transfer solutions from one system to another since various software systems report different patient data. Consequently, we limit our issue formulation to procedure and diagnostic codes, which are universally applicable regardless of their place of origin. When looking through Medicare claims data for unusual provider payments, the two- step Multivariate Outlier Detection approach comes in handy. To begin, we construct a multivariate regression model using thirteen carefully selected characteristics, and then we use those models to find the residuals. The last step is to load a generalized univariate probability model with the residuals. In order to find potential outliers in the claim data, they used probabilistic programming techniques in Stan. On the inpatient dataset collected from CMS, our experimental findings demonstrate that MCC + LSTM achieves recall scores of 50%, precision scores of 61%, and accuracy values of 59%. The accuracy, precision, and recall scores on the outpatient dataset are 78%, 83%, and 72%, respectively. We anticipate that this new study on detecting false claims with little but conclusive evidence will be initiated by the suggested issue formulation, representation learning, and solution.

2.LITERATURE SURVEY

1. Health insurance claim fraud detection using machine learning algorithms

Authors: K. Ashesh, P. Sritha, D. Vineela, and P. Swathi

Data mining and artificial intelligence processes for criminologist job social insurance cheaters are discussed in the test paper, which was drained by the connected work. In order to tackle any problems, information preparation, being an eager examination arena, provides software engineering methods and applied math investigations. A lot of associations used to be worried about collecting the right data, but now that they have information preparation, that problem is solved. False and severe protection practices might be a major problem in many countries. Looking at the data on the far side of the management level is the greatest method to find misrepresentation in a reasonable manner. This study&#39;s overarching goal is to use application behavior assessment to successfully define applications and differentiate/identify exception apps. In the same way that different criteria are used to determine if an app is accessible to a client&#39;s local content without the client&#39;s knowledge, exception applications are used to determine whether a relevant mechanical man application continues based on itsrepresentation on the Google Play Store.

2. Insurance Fraud Claims Detection using Machine Learning

Authors: Hritik Kalra, Dr. T. Senthil Kumar, and Ranvir Singh

In order to analyze insurance claims effectively, Rama Devi Burri examined several machine learning and statistical methods. Along with the difficulties encountered in using machine learning techniques, they also discussed several applications of these methods in the insurance sector. A project that will identify automotive industry fraud was suggested by Sunita Mall. In order to identify fraud triggers and determine the likelihood of claims being approved or refused, this study employs a number of statistical methods, including Logistic Regression. The goal of Pinak Patel&#39;s rule-based pattern mining is to detect and quantify healthcare fraud. The provided data shows the insurance claims for fraud as outliers according to statistical decision criteria, k-means clustering, and association rule- based mining using a gaussian distribution.

3. Using Machine Learning to Detect Insurance Claims Fraud

Authors: Abdul Rahman Alrais, Arif

Identifying and Preventing Fraud: The goal of a fraud detection system is to catch questionable actions before they reach the main system (Aisha Abdallah, 2016). Manually combing through a subset of actual fraud data to spot and identify such actions was the previous standard. The procedure has taken a long time and has been fraught with mistakes, misunderstandings, and forgotten facts due to human error. As a result, fraud detection systems have developed to automate the process and eliminate human error at the operational level of the system. However, previous iterations of these systems lacked adequate data mining techniques, which are now available in much improved forms that can yield more accurate results.

Machine learning for the diagnosis of heart failure reasons: One further area where data science and machine learning have found use is in the diagnosis of heart failure. Numerous relevant publications have been located to substantiate this study. Using big data analytics, this research aims to determine what causes heart failure. A large body of recent research indicates that big data analytics has contributed to health improvement.

The death toll from heart disease is steadily rising. Research using data mining methods to discover, predict, and ultimately cure a problem has been conducted in recent years, made possible by the availability of massive quantities of data. One study postulated that data mining is common practice due to the data deluge and the possibility of benefits from data transformation into useful insights. The authors of the 2019 article &quot;Predictive Analysis on Heart Disease Using Different Machine Learning Techniques&quot; are Niraj and R. Data mining proved to be a useful tool for them when analyzing massive datasets. When it comes to analysis, these prediction methods are useful for professionals. They made use of a dataset of cardiac patients that was accessible via the machine learning datasets at UCI. Adaptive boosting, Logistic Regression, Decision Trees, SVM, Random Forest, and K-Nearest Neighbor were among the machine learning methods used. Applying these algorithms to data may help them discover results that might aid doctors in estimating the risk of heart failure in various age groups.

4. Using Data Mining Techniques to Detect Insurance Claims Fraud

Authors: Authors: Yash Maske, Siddharth Mal, and Pinak Patel Improvements to people&#39;s health are a steady side effect of scientific and technological progress in the US. To keep up with the skyrocketing prices of high- quality healthcare, programs like Medicare are essential. The capacity of Medicare to adequately meet the healthcare requirements of eligible persons, including the elderly, is jeopardized when dishonest individuals commit fraud for malicious purposes and financial benefit. Several &quot;Big Data&quot; datasets for various Medicare programs were provided by the

Centers for Medicare and Medicaid Services (CMS) in an effort to reduce fraudulent activity. The identification of Medicare fraud is the primary subject of this article. We construct and evaluate three learners on each dataset as part of our exploratory investigation on Medicare fraud detection. The combined dataset with the Logistic Regression (LR) learner produced the greatest overall score of 0.816 on the Area under the Receiver Operating Characteristic (ROC) Curve performance measure. Part B dataset with LR came in second with 0.805. Without a statistically significant difference between the two datasets, the Combined and Part B datasets outperformed all other learners when it came to fraud detection. Based on our findings and the assumption that it is impossible to predict which Medicare program a doctor will defraud, we propose using the combined dataset to identify instances of fraudulent behavior when a doctor has submitted payments through any of the Medicare programs we looked at.

5. Insurance Claim Fraud Analysis and Detection using Machine Learning

Authors: Raj Akansh, Raj Aditya, Narayan Sharma Aditya, Kumar Saket,

Kaushik Rohit, and Rani Shallu

The term &quot;metric capacity unit&quot; is often used to describe machine learning. Computers having the implicit capacity to be learned without explicit programming are included in the research of machine learning. Processing data in the context of machine learning has come a long way in the last few years. Data mining is all about analyzing all the data that is collected. Information processing also tries to discover practical patterns within it.In contrast, processing applications like machine learning employ knowledge to find patterns in data and improve program behaviors, there by gaining a better grasp of the environment. The task of inferring meaning from labels on training data is essential to supervised machine learning. Before trying to build a short-lived perform, a supervised learning rule completes a foundational job using the sample data. As a result, it plots fresh input vectors.

3.EXISTING SYSTEM

The term &quot;metric capacity unit&quot; is often used to describe machine learning. Computers having the implicit capacity to be learned without explicit programming are included in the research of machine learning. Computer programs with sufficient alterable capability, which were previously unprotected to new knowledge, are the focus of this capacity unit&#39;s growth efforts. Supervised learning, unsupervised learning, and reinforcement learning are the three primary categories into which metric capacity unit algorithms fall. Processing data in the context of machine learning has come a long way in the last few years. Data mining is all about analyzing all the data that is collected. Information processing also tries to discover practical patterns within it. In contrast, processing applications like machine learning employ knowledge to find patterns in data and improve program behaviors, thereby gaining a better grasp of the environment. The task of inferring meaning from labels on training data is essential to supervised

machine learning. A collection of coaching samples makes up the coaching data. Every instance in supervised learning starts with a base that has an input object (the vector) and an output value (the model&#39;s execution indication) in the same way. Before trying to build a short-lived perform, a supervised learning rule completes a foundational job using the sample data. As a result, it plots fresh input vectors. There is a wide variety of applications that make use of supervised learning techniques. There is a supervised learning rule that aims to reduce knowledge to enclosed objects in a very good way, and the optimal setting gives the rule the possibility to correctly mark the class labels for near instances.

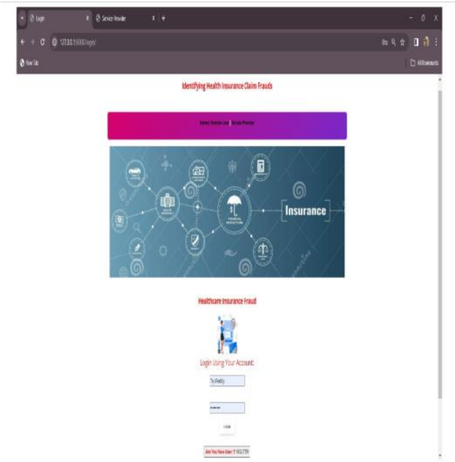
Disadvantages:

• CNMFS-based Supervised Spammer Detection with Social Interaction (CNMFSD) is not applied to the system.

• There is no ML classifier in the system that can be used to train or test datasets.

4.OUTPUT SCREENS

Remote User: Login



Identifying Health Insurance Claim Frauds

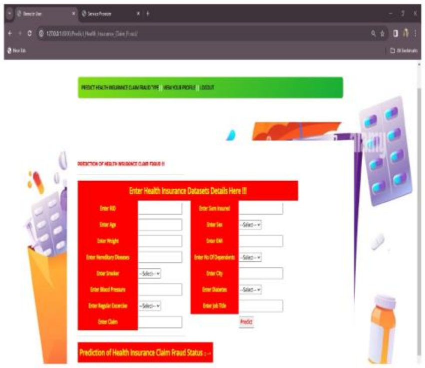
Healthcare Insurance Fraud

Login Using Your Account:

TejReddy

Are You New User !! REGISTER

Remote User: Predict Health Insurance Claim Fraud Type



PREDICT HEALTH INSURANCE CLAIM FRAUD TYPE || VIEW YOUR PROFILE || LOGOUT

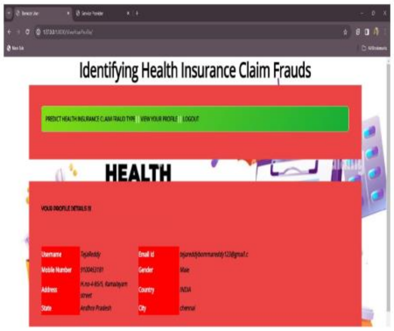
DETECTION OF HEALTH INSURANCE CLAIM FRAUD !!!

Enter Health Insurance Datasets Details Here !!!

Enter RID		Enter Sum Insured	
Enter Age		Enter Sex	--Select--
Enter Weight		Enter BMI	
Enter Hereditary Diseases		Enter No Of Dependents	--Select--
Enter Smoker	--Select--	Enter City	
Enter Blood Pressure		Enter Diabetes	--Select--
Enter Regular Excercise	--Select--	Enter Job Title	
Enter Claim			Predict

Prediction of Health Insurance Claim Fraud Status :--

Remote User: Profile



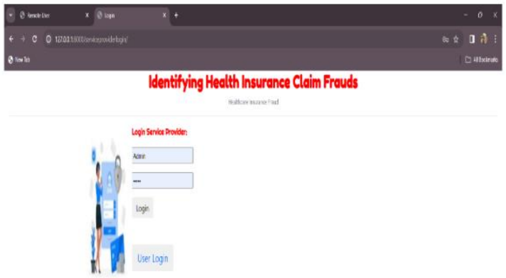
Identifying Health Insurance Claim Frauds

PREDICT HEALTH INSURANCE CLAIM FRAUD TYPE || VIEW YOUR PROFILE || LOGOUT

YOUR PROFILE DETAILS !!!

Username	TejaReddy	Email Id	tejareddybommareddy123@gmail.c
Mobile Number	9100453187	Gender	Male
Address	H.no-4-85/5, Ramalayam street	Country	INDIA
State	Andhra Pradesh	City	chennai

Service Provider: Login



Identifying Health Insurance Claim Frauds

Healthcare Insurance Fraud

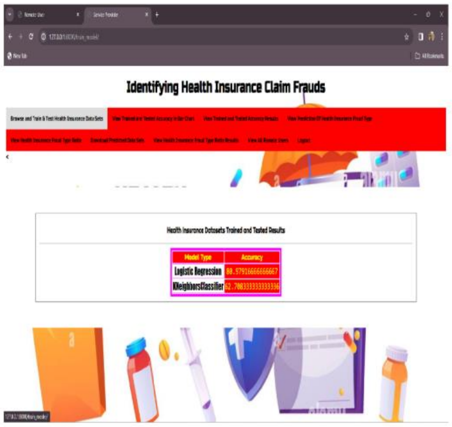
Login Service Provider:

Admin

Login

User Login

Service Provider: Train &amp; Test Health Insurance Dataset



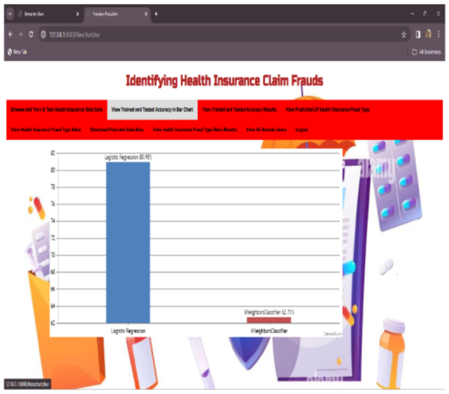
Identifying Health Insurance Claim Frauds

Browse and Train & Test Health Insurance Data Sets | View Trained and Tested Accuracy in Bar Chart | View Trained and Tested Accuracy Results | View Prediction Of Health Insurance Fraud Type | View Health Insurance Fraud Type Ratio | Download Predicted Data Sets | View Health Insurance Fraud Type Ratio Results | View All Remote Users | Logout

Health Insurance Datasets Trained and Tested Results

Model Type	Accuracy
Logistic Regression	88.97916666666667
KNeighborsClassifier	62.708333333333336

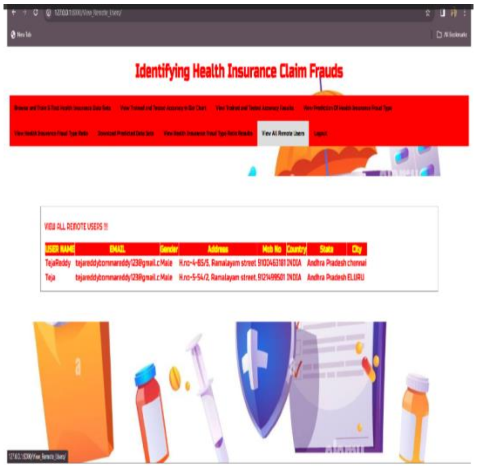
Service Provider: Trained Accuracy in Bar Chart



Identifying Health Insurance Claim Frauds

Browse and Train & Test Health Insurance Data Sets | View Trained and Tested Accuracy in Bar Chart | View Trained and Tested Accuracy Results | View Prediction Of Health Insurance Fraud Type | View Health Insurance Fraud Type Ratio | Download Predicted Data Sets | View Health Insurance Fraud Type Ratio Results | View All Remote Users | Logout

Service Provider: View Remote Users


127.0.0.1:8000/View_Remote_Users/
New Tab
All Bookmarks
Identifying Health Insurance Claim Frauds
Browse and Train & Test Health Insurance Data Sets
View Trained and Tested Accuracy in Bar Chart
View Trained and Tested Accuracy Results
View Prediction Of Health Insurance Fraud Type
View Health Insurance Fraud Type Ratio
Download Predicted Data Sets
View Health Insurance Fraud Type Ratio Results
View All Remote Users
Logout

VIEW ALL REMOTE USERS !!!

USER NAME	EMAIL	Gender	Address	Mob No	Country	State	City
TejaReddy	tejareddybommareddy123@gmail.c	Male	H.no-4-85/5, Ramalayam street	9100463181	INDIA	Andhra Pradesh	chennai
Teja	tejareddybommareddy123@gmail.c	Male	H.no-5-54/2, Ramalayam street,	9121499501	INDIA	Andhra Pradesh	ELURU

5.CONCLUSION

As a feature generation and classification method, we present the issue of fraudulent insurance claim identification in this study. Due to legal restrictions and software system discrepancies, we construct the issue over minimum, definite claim data consisting of procedure and diagnostic codes rather than broader datasets. As a novel representation learning strategy, we provide clinical ideas as an alternative to procedure and diagnostic codes. Our findings show that there is room for improvement in the ability to detect fraudulent healthcare claims using limited data. For different idea sizes and replacement probability, both MCC and MCC + RPCA behave consistently throughout the negative claim generation process. In the inpatient dataset, MCC + LSTM achieves recall of 50%, precision of 61%, and accuracy of 59%. Additionally, on the outpatient dataset, it displays recall scores of 72 percent, accuracy of 78 percent, and precision of 83 percent. New study on detecting fraudulent insurance claims utilizing minimum but conclusive evidence will be initiated by the suggested issue formulation, representation learning, and solution.

6.REFERENCES

1] National Health Care Anti-Fraud Association, "The challenge of health care fraud," https://www.nhcaa.org/resources/health-care-antifraud-resources/the-challenge-of-health-care-fraud.aspx, 2020, accessed January, 2020.

[2] Font Awesome, "Image generated by free icons," https://fontawesome.com/license/free, 2020, online.

[3] National Health Care Anti-Fraud Association, "Consumer info and action," https://www.nhcaa.org/resources/health-care-anti-fraudresources/consumer-info-action.aspx, 2020, accessed January, 2020.

[4] W. J. Rudman, J. S. Eberhardt, W. Pierce, and S. Hart-Hester, "Healthcare fraud and abuse," Perspectives in Health Information Management/AHIMA, American Health Information Management Association, vol. 6, no. Fall, 2009.

[5] M. Kirlidog and C. Asuk, "A fraud detection approach with data mining in health insurance," Procedia-Social and Behavioral Sciences, vol. 62, pp. 989–994, 2012.

[6] V. Rawte and G. Anuradha, "Fraud detection in health insurance using data mining techniques," in 2015 International Conference on Communication, Information &amp; Computing Technology (ICCICT). IEEE, 2015, pp. 1–5.

[7] C. Phua, D. Alahakoon, and V. Lee, "Minority report in fraud detection: classification of skewed data," Acm sigkdd explorations newsletter, vol. 6, no. 1, pp. 50–59, 2004.

[8] T. Ekina, F. Leva, F. Ruggeri, and R. Soyer, "Application of bayesian methods in detection of healthcare fraud," chemical engineering Transaction, vol. 33, 2013.

[9] J. Li, K.-Y. Huang, J. Jin, and J. Shi, "A survey on statistical methods for health care fraud detection," Health care management science, vol. 11, no. 3, pp. 275–287, 2008.

[10] R. J. Freese, A. P. Jost, B. K. Schulte, W. A. Klindworth, and S. T. Parente, "Healthcare claims fraud, waste and abuse detection system using non-parametric statistics and probability-based scores," Jan. 19 2017, uS Patent App. 15/216,133.