SYMBOLIC DATA APPROACH FOR FRAUD DETECTION IN HEALTHCARE SYSTEM

[1]Sahana Munavalli,[2]Sanjeevakumar M. Hatture
[1] Department of Computer Science and Engineering, Basaveshwar Engineering College(Autonomous), Bagalkot - 587103, Karnataka State, India.
Sahana.s.munavalli@gmail.com
[2]Department of Computer Science and Engineering, Basaveshwar Engineering College(Autonomous), Bagalkot - 587103, Karnataka State, India.
smhatture@gmail.com

***Abstract*—** Fraud and abuse in the health care system have become a significant concern and that too inside health insurance organizations. In the last decade due to expanding misfortunes in incomes, handling medical claims have become a debilitating manual assignment, which is done by a couple of clinical specialists who have the duty of endorsing, adjusting, or dismissing the appropriations mentioned inside a restricted period from their gathering. Standard data mining techniques at this point do not sufficiently address the intricacy of the world. In this way, utilizing Symbolic Data Analysis is another sort of data analysis that permits us to address the intricacy of the real world and to recognize fraud in the healthcare data. In the era of digitization, the frauds are found in all categories of health insurance. It is finished next to deliberate trickiness or fraud for acquiring some pitiful advantage in the form of health expenditures. Bigdata analysis can be utilized to recognize fraud in large sets of insurance claim data. In light of a couple of cases that are known or suspected to be false, the fraud detection technique computes the closeness of each record to be fake by investigating the previous insurance claims. The investigators would then be able to have a nearer examination for the cases that have been set apart by data mining programming. One of the issues in healthcare system claims is the abuse of the medical insurance. Manual detection of frauds in the healthcare industry is strenuous work. In the proposed work the symbolic data object representation reduces the complexity of the medical data and machine learning and symbolic data analysis recognize the fraud detection. The symbolic data approach for fraud detection in healthcare system achieves promising results and can be extended to other healthcare systems.

Keywords—Fraud detection; Anomaly detection; Data mining; Health insurance; Symbolic data

I. Introduction

Healthcare systems policies are intended to design and give clinical consideration to individuals. Clinics, Hospitals, and community health agencies are the part of the healthcare systems. These healthcare systems are unpredictable, and numerous things should be thought about kinds of emergency hospital systems, patient care, insurance, healthcare providers, and lawful issues. Medical services have become a significant consumption in the United States(US) since 1980. Both the size of the healthcare sector and the gigantic volume of cash included make it an alluring fraud target. There are various kinds of healthcare care systems, viz. National Health Service, non-profit national health system, National Health Insurance System (NHIS), social health insurance, private health system, social health insurance system, etc. According to the Office of Management and Budget, in 2010, about 9%, or around $47.9 billion of the US Medicare expenditure was lost because of fraud. Consequently, effective fraud detection is significant for diminishing the expense of the healthcare system. Fraud is the crime of acquiring cash or monetary advantages by a trick or by lying. Fraud can be spread comprehensively, and it is very exorbitant to ensure the system. The medical services industry is a multifaceted framework with various moving segments. Simultaneously, fraud in this industry is transforming into a critical issue. Distinguishing medical services fraud and misuse, in any case, needs concentrated clinical information. As indicated by the report of Association Certified Fraud Examiner Report Estimation (ACFE), Organizations in the Worldwide Lose 5% of revenues to fraud viz. projected losses exceed $3.5 trillion Worldwide. In medical services, fraud can happen in various circumstances, from superfluous and duplicate tests and strategies to hacking into a patient's clinical records to submit bogus claims. In clinical medical care systems, fraud might be forced like billing for services not rendered, billing for a non-covered assistance as a covered help, misrepresenting locations of service, misrepresenting areas of administration, misrepresenting supplier of overhauling of deductibles or potentially co-installments, incorrect revealing of diagnoses or procedures (incorporates unbundling), overutilization of services, corruption (payoffs and bribery), false or pointless issuance of doctor prescribed medications. To distinguish the misrepresentation designs different data analyzing methods are utilized subsequently they are recognized as regular examples.

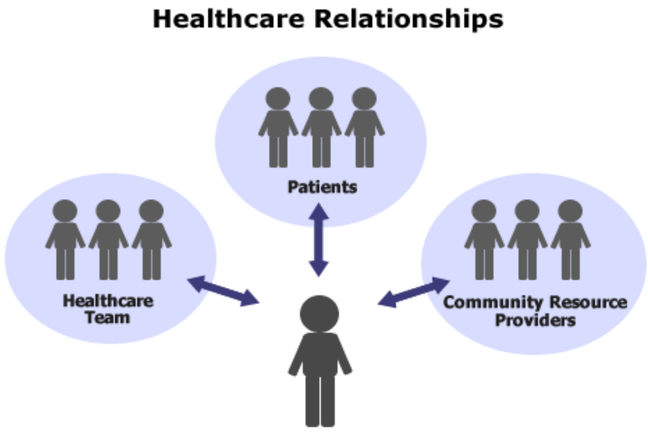

Fig 1: Healthcare System groups

Fraud is the crime of gaining money or financial benefits by a trick or by lying. Healthcare frauds can occur in many different situations, from unnecessary and duplicate tests and procedures to hacking into a patient's personal medical records to submit false claims. Figure 1 shows the different groups of the health care systems. In healthcare team, there are various types which need to be known i.e., types of care, and the roles of each member of the healthcare team. As there will be interaction of work with patients, there will be need to understand different types of insurance, how to help uninsured patients and how to protect patient rights and privacy and also there is a need to know, what community resources are available and how to access those services for patients.

The patient health information is crucial and complex as it contains blood pressure, body temperature, electrocardiogram (ECG) signals, electroencephalogram (EEG) signals, electromyography (EMG) signals, serum, potassium, calcium, creatinine level and many more details and data are in unstructured manner. Hence the symbolic object representation is employed to minimize the complexity and redundancy in the healthcare data. The major benefits on employing the healthcare data into symbolic objects are; they give a summary of the original symbolic data in an illustrative manner; they can be easily changed in terms of the query of a database; being autonomous of the independent data table they can recognize any coordinating with individual portrayed in any data table; in the utilization of their elucidating part, they can give another symbolic data table of a more elevated level on which a symbolic data analysis of the subsequent level can be applied; they can effectively join a few properties dependent on different variables coming from different arrays and different underlying populations.

With the assistance of symbolic data, fraud in health care can be distinguished by utilizing symbolic similarity measures and symbolic clustering. The symbolic clustering methodology forms composite symbolic objects items utilizing a Cartesian join operator when two symbolic objects are merged. This composite object with the rest of the object is additionally utilized for similarity analysis. The rest of the paper is organized into four sections: section 2 reviews the developments in fraud analysis in healthcare systems. Further the common types of frauds healthcare systems are enlisted in section 3. Symbolic data representation is described in section 4. The proposed model of fraud detection in healthcare systems using symbolic data analysis is described in section 5. The section 6 concludes the proposed work.

LTERATURE REVIEW

The Centers for Medicare and Medicaid Services (CMS) releases health care information which is utilized by the greater part of the scientists for healthcare fraud detection. Lucy Fricker [1] proposed enterprise risk management(ERM) can assist with diminishing sorts of fraud and used for reducing the fraud. ERM is a process, affected by an effected by an entity's board of directors the board and other work force, applied in methodology setting and across the enterprise, intended to recognize potential events that may influence the entity, and oversee hazard to be inside its danger craving, to give sensible confirmation with respect to the accomplishment of element goals of entity.

Qi Liu[3] proposed a Geo-location clustering model. In this, preliminary knowledge of healthcare system and its fraudulent behaviors, and analyzes the characteristics of health care data and they compare currently proposed fraud detection approaches like clustering, decision making, etc. with Geolocation clustering model is performed. Yi Peng [4] proposed two health insurance datasets namely SAS Enterprise Miner and CLUTO datasets and used to detect frauds involved in claiming the insurance. SAS Enterprise Miner is an advanced analytics data mining tool intended to help users quickly develop descriptive and predictive models through a streamlined data mining process. CLUTO is a software package for clustering low and high dimensional datasets and for analyzing the characteristics of the various clusters. Shivani S. Waghade[5] presented machine learning and data mining are used for healthcare fraud detection and explained types of frauds.

Healthcare fraud and also conclude that the advanced machine learning techniques and newly acquired sources of the healthcare data would be forthcoming subjects of interest to make the healthcare affordable, to improve the effectiveness of healthcare fraud detection and to bestow a top-quality on healthcare systems. Ajith Abraham,[6] proposed a classification framework for the application of data mining techniques to insurance fraud detection and classified into three types of insurance fraud (automobile insurance, crop insurance, and healthcare insurance) and six classes of data mining techniques (classification, regression, clustering, prediction, outlier detection, and visualization).The main data mining techniques used for insurance fraud detection are logistic models, Decision trees, the Naïve Bayes, and support vector machine.

Analysis. In this, proposed classification framework for the application of data mining techniques to insurance fraud Detection and the analysis results show that automobile insurance fraud in the most covered area of research (57%). The data mining application class that was used in most of the papers is the classification (57%). Matthew Herland [7] proposed a Supervised learning algorithm for class imbalanced data. Every day, there are a massive number of financial transactions generated by physicians administering healthcare services, such as hospital visits, drug prescriptions, and other medical procedures. Most of these financial transactions are conducted without any fraudulent intent, but there are a minority of physicians who maliciously defraud the system for personal gain. In machine learning, when a dataset portrays this discrepancy in class representation (i.e., a low number of actual fraud cases), it is known as a class imbalance.

The main issue attributed to class imbalance is the difficulty in discriminating useful information between classes due to the over-representation of the majority class (non-fraud) and the limited amount of information available in the minority class (fraud).UC San Diego[8] proposed a health care fraud and abuse blockchain technical framework and prototype using key blockchain tools, for secure data storage and consensus mechanisms, which make the claims adjudication process more patient-centric to identify and prevent health care fraud and abuse. The primary aims of the blockchain solution are to (1) improve detection of potentially fraudulent and illegal health care transactions and reimbursements, (2) create a more inclusive process for validating claims deploying a patient-centric approach, and (3) enhance efficiency in the claims adjudication process through smart contract automation.

The variety of frauds in the healthcare systems and insurance frauds are explored in [9,10]. Angadi S.A and Hatture S.M in [11] have proposed the graph theoretic approaches are used to identify the correlation between the features of the hand image data. In-order to manage ordered and non-ordered mixed feature-type symbolic data, a previous pre-processing step was introduced to obtain a suitable homogenization of mixed symbolic data into modal symbolic data represented by weight distributions. The dynamic cluster algorithm with adaptive distances locally optimizes an adequacy criterion that measures the fitting between the classes and their representatives (prototypes). V. Ravi [12] a hierarchical agglomerative symbolic clustering methodology is Proposed. The procedure is based on the physical phenomenon in which a system of particles in space converges to the centroid of the system due to gravitational attraction between the particles are merged.

The procedure terminates at some stage where there no available for merging. Anderson F.B.F. Costa [13] The main idea of this kernel k-means (KKM) consists of manipulating a kernel function for interval data to compute the distance between two vectors of interval data. kernel k-means(KKM) method is an extension of the classic K-means Kernel to symbolic interval data. A new clustering method is proposed for symbolic interval data based on kernel. This method is an extension of the classic K-means Kernel to symbolic interval data. The evaluation of this method can be done by comparing with a dynamic clustering method for interval data having adequacy criterion-based (on adaptive for each cluster) Euclidean distances carried out and result provided by clustering methods were assessed correctly using rand index by considering the synthetic and applications of real data sets. K. Chidananda Gowda [14] the clustering methodology forms composite symbolic objects using a Cartesian join operator when two symbolic objects are merged.

Merging is the process of gathering, based on a similarity measure two samples and assigning them same cluster. The clustering methodology is proposed in a composite object when two selected objects are merged. This composite object, along with the rest of the objects of the set, is used in further similarity analysis. similarity components due to “position” which feature type is Quantitative and “span”,” content” which feature type is qualitative [15]. A composite symbolic object is a new object resulting from merging two symbolic objects using a Cartesian join Operatorial. Vipin Kumar [16] explained a variety of similarity measures. similarity for continuous data is relatively well-understood, but for categorical data, the similarity computation is not straightforward, so for this data-driven similarity measures have been proposed for categorical data.

In data-driven similarity measures for categorical data, a key task is to identify the characteristics of a categorical data set that affect the behavior of a similarity measure such as the size of data, number of attributes, number of values taken by each attribute, etc. This work used outlier detection as the underlying data mining task for the comparative evaluation. It will be useful to know if the relative performance of these similarity measures remains the same for the other data mining tasks. J. D. Kittoe [17] In this, the researcher focused on malaria cases data. In tackling the issue of malaria in a more cost-effective means, patterns were explored in finding the mean cost of drugs for the treatment of malaria. One of the major objectives is, to use data mining techniques to detect fraud and abuse in the NHIS concerning malaria-related cases[18, 19]. Some of the issues and challenges are identified in the literature and enlisted in the following.

- Fetching irrelevant data from the data set.
- Categorizing large data set using classical data approach.
- Not easy to understand data when the dataset is more in classical data.
- Identification of genuine data.
- Using pattern recognition with the help of symbolic data.

Hence there is a scope to represent the complex healthcare data[20, 21] containing the numeric, graphical, and video and signal data, by employing symbolic data representation. Using Symbolic Data, the healthcare data can be easily segregated in range which will help in detecting the fraud and Symbolic data analysis gives a new way of extending the standard input to a set of classes of individual entities. Such classes often represent the real units of interest, using the symbolic data analysis more efficiently the output data can be retrieved.

FRAUD TYPES IN HEALTH CARE SYSTEM

Fraud can occur in multiple situations, however, to handle the circumstance or to deal with the situation, the board or the associations should think about the intercession of innovation to keep away from the maltreatment of the assets. A portion of the misrepresentation types in the medical coverage area in India are as per the following:

- Charging over the top costs for treatment or medication.
- The surprisingly maximum number of solicitations for a specific insured in the brief period of time (3-4 days).
- Insurance transaction(s) where the insured has got some treatment or medication yet either has not paid any portions or has paid just the first installment.
- Cases where the insured purchasing medication without a clinical assessment.
- Claiming clinical solicitations with dates previously or after the start of the insurance time frame (this is allowed at times).
- An extreme number of clinical cases in a particular period.
- Bank account number changes of a colleague like office, health center, or drug store.
- Excessive quantities of manual receipt request whose sums are more modest than the standard assessment limit. claims whose payable sums are more noteworthy than the receipt sum that the insurance agency will pay.

Greater part of fraud in the clinical industry has a predetermined number of examples that are normally known to the insurance experts. In medical coverage can be explicit to every nation exploiting the deficiency of the applicable enactment or being influenced by the neighborhood culture. For instance, individuals in nations with a collectivist culture may have a higher inclination to mishandle the framework contrasted with the nations with individualistic societies. The fraud may occur in any manner. In-order to alleviate the issues and challenges posed by various frauds in healthcare systems, the symbolic data is employed.

SYMBOLIC OBJECT

Representative items are described by rational blend of occasions leaning toward qualities and factors in which the factors can take at least one quality need not be portrayed on similar factors. Under a nonformal depiction of Symbolic objects of the kind of Assertion, Hoard and Synthetic are represented.

As the data ends up being fuzzier than the standard ones since they contain internal assortment and are coordinated. To summarize colossal plans of information, the need to loosen up standard data assessment methods to representative information investigation is growing to get more accurate information and summarize wide broad dataset. Further with the assistance of approaches present in symbolic data analysis (SDA), can investigate, and handle the information. The motivation behind SDA is to expand information mining procedures and conventional measurements to more elevated level units.

Representative information examination gives another perspective in data science by stretching out the standard contribution to a bunch of classes of individual elements. Such classes frequently address the genuine units of interest. To consider fluctuation between the individuals from each class, classes are portrayed by stretches, dispersions, set of classifications or numbers once in a while weighted, and such. In that manner, to get new sorts of information, called 'representative' as they cannot be decreased to numbers without losing a lot of data. The representative information table is assembled where the columns are classes, and the factors can take emblematic qualities. Further concentrate new information from these new sorts of information by somewhere around an expansion of computer statistics and data mining to emblematic information. In Symbolic data, the information is not limited as traditional information (standard information) and the information can be addressed by records, stretches, range, and so on as displayed in Table 1.

Sl.No	Age Range	Blood Pressure (mm/Hg)	City	Type of Cancer	Gender
1	25,35	76/125	Bihar	Brain tumor	Male
2	57,67	95/135	Goa	Lung, liver	Male
3	40,55	85/135	Kerala	Prostate	Male
4	37,47	87/126	Delhi	Breast p, lung(1-p)	Female

Table 1:Symbolic Data Representation

Assertion Objects

An assertion object is a conjunction of events pertaining to a particular object. Following is an example for an Assertion object:

a = [color={red, yellow}]&[size = 27]&[price = [l0 – 150]. Here a is an Assertion object, having the following properties:

1) color is either red or yellow 2) size equals27 3) price ranges between 10 and 150.

Hoard Objects

A hoard object is a conjunction of two or more Assertion objects

h =[Display(PC1)=color] & [RAM(PC1)=8GB] & [Keys(PC1) = [51-69]] & [Display(PC1=GREY] & [RAM(PC2)=6GB] & [Keys(PC2) = [62-71]].

It means that the Hoard object h consists of two elementary objects:

1) "PC1" having a color Display, 8GB RAM, and Keys

2) "PC2" having B&W Display, 6GB RAM, and Keys between 51 and 69. between 62 and71.

Synthetic Objects

Following is an example for a Synthetic object: A Synthetic object is a conjunction of two or more Hoard objects

s = hl & h2 =[type(Q1)=Expressway] & [vehicles(Q1)=2] [type(Vl)=car] [type(V2)=truck] [type(V3)=bus] & [type(Q2)=Main Road]& [vehicles(Q2)=1] & & [color(Tl)=blue] & [rnoving(El)=rl] & & [color(E2)=red] & [moving(E2)=rl] && [color(E3)=green] & [moving(E3)=r2].

METHODOLOGY

The graph proposed of fraud recognition in medical care framework is portrayed in Figure 1. Initially, input should be given as a dataset, which comprises of complex information, to this information cleaning and preprocessing steps to be completed in light of the fact that to consider how precisely the missing information to be filled. In the event that scaling of highlight is thought of and how it should be finished. Here there will be a requirement for Dummy factors or a genuine dataset? Is information going to be encoded? Regardless of whether encoding of sham factors is finished? after pre-preparing step, change of the information dataset into representative information to be done and investigation of it utilizing the SDA method ought to likewise be finished. At long last, by utilizing AI calculations like bunching/dynamic arrangement of information to be done where the dataset will be veritable or not.

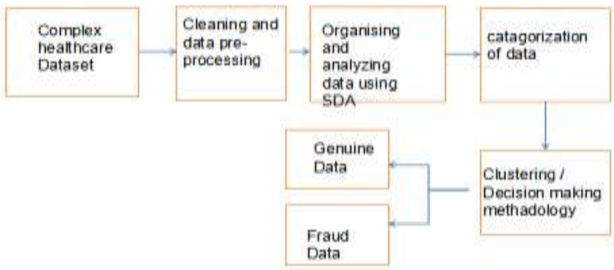

Figure 1:Fraud Analysis Using SDA Workflow

The graph of the proposed misrepresentation identification in medical services framework utilizing representative information examination is portrayed in Figure 1. The distinctive handling phases of the proposed procedures are procurement of medical services information (Dataset), information cleaning and pre-preparing, arranging, and addressing the information into symbolic item and dissecting information utilizing representative information examination, arrangement of the Symbolic information lastly settling on the choice of information related with extortion or veritable utilizing design coordinating with strategy. The usefulness of individual square is investigated in the accompanying:

A. Health Care Data Set:

The patient information is gathered from the patient information gathered may contain the traditional information put away in a data set. To address the traditional information is gone through the pre-handling stage for additional element extraction. viz. information obtaining and development of the medical services Data Set.

B. Cleaning and Preprocessing:

The gathered information into emblematic item the information is gathered, regularly the information will not be in a structure that is reasonable for handling. To make the information appropriate for handling and addressed into representative article it is fundamental to change them into a configuration the information, for example, multidimensional, time series, or semi-organized organization. The element extraction stage is frequently acted in corresponding with information cleaning in which loud information and immaterial information are taken out from the assortment i.e., in gathered information some information might be missed, or some might be insignificant, for that reason information cleaning is to be finished. The consequence of this stage is an all-around organized informational index, which can be adequately utilized for portrayal as emblematic article.

C. Organizing and Analyze data:

Once the dataset is cleaned and pre-prepared, the following stage is to break down the information. With the assistance of Symbolic Data Analysis, utilizing the referenced. Procedure huge and complex information can be prepared and pictured by changing over the intricate information into the scope of information (Symbolic Object).

D. Pattern Extraction and Matching:

Subsequent to investigating the information through emblematic information examination, the yield viz. emblematic article will be sorted as Genuine or Fraud dependent on the examples perceived (progressive grouping calculation). To recognize, if the given dataset is veritable, with the assistance of dubious examples with AI calculations utilizing emblematic information is performed. The exploratory outcomes are promising and identifying with a precision of 99%. The emblematic information portrayal of the medical care information and representative information examination are all around familiar for the medical services frameworks.

RESULTS AND DISCUSSION

Once the data is trained and tested, based on the considered data, graphical result is shown below, the graphical representation contains the result with True positive and False positive rate with respect to the data input after comparison. For the segregation, the data is tested and trained using machine learning algorithms like naive bayes for heart disease dataset and k-means for skin cancer dataset. At the time of train and test, it will take random values and calculate the accuracy, it has to compare with the user data and give the output as legitimate or fraud.

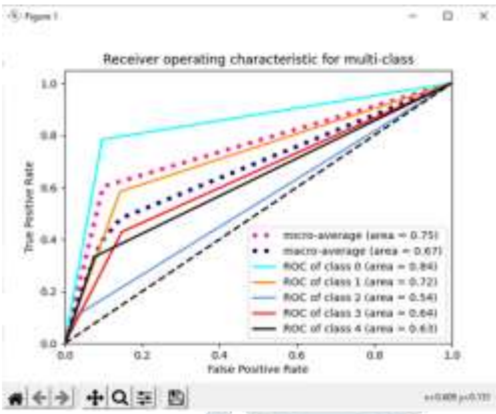

Figure 1

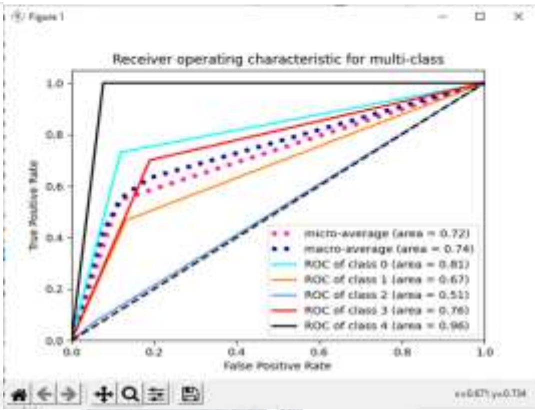

Figure 2

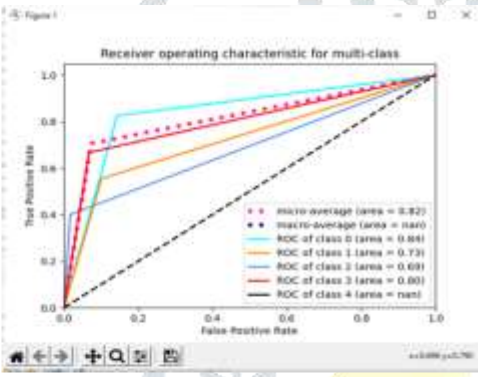

Figure 3

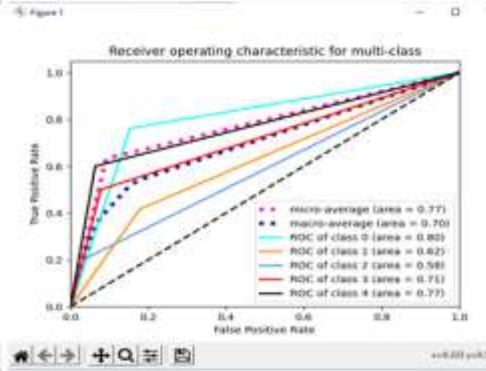

Figure 4

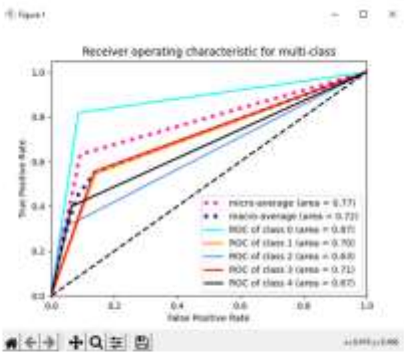

Figure 5

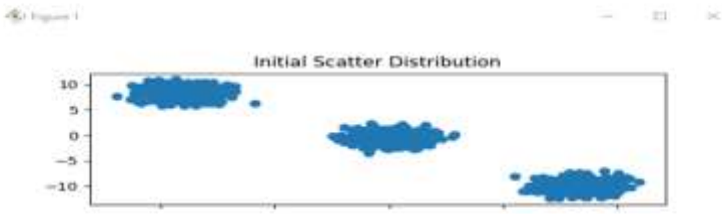

Figure 6

With respect to above figures, i.e., from Figure1-5 are the output graphs for the Heart disease Fraud Detection, where the graph is shown as a class diagram output with the true and false positive rate. Referring to the Figure 6, for the Skin cancer fraud detection, the graph k-means cluster is shown with the scattered values and depending upon the similar grouping of symptoms.

The test results are added underneath for the both the tests conveyed for Skin Cancer Fraud Detection and Heart Disease Fraud Detection. For Skin Cancer, content comparability is being thought of and comparatively for coronary illness, length closeness is thought of.

For Skin Cancer Fraud Detection K-implies calculation is being utilized. It is an iterative calculation isolates dataset into independent bunches dependent on the similitude, by this way that each dataset will be isolated with comparable gatherings that has comparable properties.

Attributes considered for Skin Care Fraud Detection:

1. SO_{skin_cancer}=Melanoma{one half is unlike other half,irregular,6mm or larger, varied from area to other, Changing color, size, shape}
2. SO_{skin_cancer}=Cancer{Asymmetry,Border,Diameter,Color,Evolution}
3. SO_{skin_cancer}=Squamous_cell_carcinoma{one half is like other half,scalloped,6mm or larger, no variation, Changing color}
4. SO_{skin_cancer}=Bascal_cell_carcinoma{one half is unlike other half, irregular, 6mm or lager, varied from area to other, Changing color, shape, size}
5. SO_{skin_cancer}=Merkel_cell_Cancer{one half is unlike other half, Poorly Defined,6mm or larger, varied from other area, Changing shape}

Skin Cancer Types:	S0=Melanoma, S1=Squamous cell carcinoma, S2=Bascal cell carcinoma, S3=Markel cell Cancer
Asymmetry:	A0= One half is unlike another half, A1= One half is like another half
Border:	B0= Irregular, B1= Not clear, B2= Scalloped, B3=Poorly Defined
Color:	C0= Varied from Area to other, C1=No Variation, C3= Other
Diameter:	D0=6mm or larger, D1=Below 6mm, D2=Other
Evolution:	E0= Change in color, size, shape, E1= Change in color, E2=Change in Shape
Status	G= Seems to be Genuine, F= Seems to be Fraud

Table 2: Abbreviations

In the below mentioned Table 3, Table 2 abbreviations are utilized to show the output data.

Sl. No	Skin cancer types	Asymmetry	Border	Color	Diameter	Evolution	Status
1)	S0	A0	B0	C0	D0	E0	G
2)	S0	A1	B1	C0	D0	E0	F
3)	S1	A1	B2	C1	D0	E1	G
4)	S1	A1	B1	C1	D0	E1	F
5)	S2	A0	B0	C0	D0	E0	G
6)	S2	A1	B0	C0	D0	E0	F
7)	S3	A0	B3	C0	D0	E2	G
8)	S3	A0	B3	C0	D0	E1	F

Table 3: Output Data

For Heart Disease Fraud Detection, Naive Bayes calculation is utilized which is a directed learning calculation, it depends on Bayes hypothesis and utilized for tackling characterization issues. It is for the most part utilized in arrangement of text that has a high-dimensional preparing dataset. It is a probabilistic classifier, i.e., it predicts based on the likelihood of an article.

Considering the underneath table 2, the contractions of the phrasings are added regarding the particular Cancer types, and the different properties which are being considered during the execution. The information given by the client through Excel document as a dataset will be passed and the outcome will be acquired will be founded on the separate calculation utilized.

Attributes considered for Heart Disease Fraud Detection:

SOHeart_Disease=Heart{Age, Gender, Cp, restbp, chol, fbps, restecg, thalac ,exang, slope, oldpeak, ca, thal}

The above-mentioned respective attributes are considered while comparing the fraud detection.

Age:	A1=Above 15 and below 25, A2= Above 35 and below 45 A3= Above 45 and below 65, A4= Above 65 and below 100
Gender:	G1= 1 for Male, G2= 0 for Female
Chest Pain:	1= Typical Angina, 2= A-Typical Angina 3= non-Angina, 4=A-symptomatic pore
Rest BP:	R1= Normal (<120/80 mm Hg), R2=Prehypertension(120-129 and 80-89-mm Hg) R3= Hypertension(>130/90 mm Hg)
Cholesterol:	C1= Less than 200 mg/dL, C2=200-239 mg/dL C3=>240 mg/dL
ECG:	E1= 0 for Normal, E2= 1 for Having ST-T E3=2 for Hyper Therapy
Status	G= Seems to be Genuine, F= Seems to be Fraud

Table 4: Abbreviations

In the below mentioned Table 5, Table 4 abbreviations are utilized to show the output data.

Sl. No	Age	Gender	Chest pain	Rest BP	Cholesterol	ECG	Status
1)	A0	G0	2	R1	C1	E0	G
2)	A0	G0	2	R2	C0	E0	F
3)	A1	G0	2	R2	C1	E2	G
4)	A1	G0	2	R0	C0	E0	F
5)	A2	G1	3	R2	C2	E1	G
6)	A2	G1	1	R0	C0	E0	F
7)	A3	G1	1	R2	C1	E2	G
8)	A3	G1	4	R0	C0	E0	F

Table 5: Output Data

CONCLUSION

Abnormality discovery, clustering., and arrangement can effectively recognize irregularities or exceptions in enormous arrangements of data. This can be helpful for the insurance business which has issues with false cases. When the bizarre cases are identified, a few examinations should be made on them to direct a careful examination. The principal task in these examinations is to limit the objective for identifying fakes, the developed software will help in detecting the fraud based on the processed data i.e., trained, and tested data which will compare with the real-time user input for providing the precise results of false cases and genuine cases.

REFERENCES

[1] Lucy Fricker, "Causes and Challenges of Healthcare Fraud in the US", North Carolina State University Raleigh, NC, USA,2013.

[2]https://www.researchgate.net/publication/220924701

[3]Qi Liu," Healthcare fraud detection: A survey and a clustering model incorporating Geo-location information", Rutgers University Newark, New Jersey, United States,2013.

[4]Yi Peng," Application of Clustering Methods to Health Insurance Fraud Detection", Institute of Information Science, Technology & Engineering, USA,2006.

[5]Shivani S. Waghade," A Comprehensive Study of Healthcare Fraud Detection based on Machine Learning", Shri Ramdeobaba College of Engineering and Management, Nagpur,2018.

[6]Ajith Abraham," Computational Intelligence Models for Insurance Fraud Detection", Machine Intelligence Research Lab, WA, USA,2013.

[7]Matthew Herland, "The effects of a class rarity on the evaluation of supervised healthcare fraud detection models", Florida Atlantic University,777 Glades Road, Boca Raton, FL, USA,2019.

[8]UC San Diego," Combating Health Care Fraud and Abuse: Conceptualization and Prototyping Study of a Blockchain Antifraud Framework", School of Medicine, Department of Anesthesiology and Division of Infectious Diseases and Global Public Health, La Jolla, CA, United States,2020.

[9]P. Naga Jyothi," Performance on Fraud Detection in Medical Claims of Healthcare Data",2019.

[10]Ms. Meena Kumari," FICCI Working Paper on Health Insurance Fraud", Joint Director, IRDA,2019.

[11] Shanmukhappa A Angadi, Sanjeevakumar M Hatture, "Hand geometry based user identification using minimal edge connected hand image graph", IET Computer Vision, Vol. 12, Issue 5, pp. 744-752, 2018.

[12] Renata M. C. R. de Souza, "A Clustering Method for Mixed Feature-Type Symbolic Data using Adaptive Squared Euclidean Distances", Centro de Informatica - CIn / UFPEAv. Prof. Luiz Freire, s/n - Cidade Universitaria,2007.

[13] T. V. Ravi, "Clustering of Symbolic Objects Using Gravitational Approach", IBM Solutions Research Center, Indian Institute of Technology, New Delhi, India, 1999.

[14]AndersonF.B.F. Costa," A Kernel K-means Clustering Method for Symbolic Interval Data", Federal University,2010.

[15] Shanmukhappa A Angadi, Sanjeevakumar M Hatture, Face Recognition Through Symbolic Modeling of Face Graphs and Texture, International Journal of Pattern Recognition and Artificial Intelligence, 33(12), 2019.

[16]Miin-Shen Yang, "Fuzzy clustering algorithms for mixed feature variables", Department of Applied Mathematics,2003.

[17] Vipin Kumar "Similarity Measures for Categorical Data: A Comparative Evaluation", Proceedings of the SIAM International Conference on Department of Computer Science and Engineering University of Minnesota,2008

[18]J. D. Kittoe, "Exploring fraud and abuse in National Health Insurance Scheme (NHIS) using data mining technique as a statistical model", Department of Computer Science and Engineering,2017.

[19] K. Chidananda Gowda," Symbolic Clustering Using a New Similarity Measure", S.J. Coll. of Eng., Mysore, India,1992.

[20]Rashmi. P.Karchi Nagarajan Munusamy, Exploration of Unmixing and Classification of Hyperspectral Imagery, International Journal of Innovative Technology and Exploring Engineering(IJITEE), 8(7), pp. 723-733, 2019.

[21]Sanjeevakumar M Hatture, NagaveniKadakol, Clinical diagnostic systems based on machine learning and deep learning, Demystifying Big Data, Machine Learning, and Deep Learning for Healthcare Analytics, pp. 159-183, 2021.