Monetary Mine Guard : Detecting Upcoming Fraud using Data Mining Techniques

Mr.Bhanu prasad Gorantla[1] ,Harsha Vardhini[2],Madhuri[3],Praveena[4]

[1]Assistant Professor, Department of CSE, Malla Reddy Engineering College for Women, Hyderabad, bhanu.gorantla2020@gmail.com

[2,3,4]UG Students, Department of CSE, Malla Reddy Engineering College for Women, Hyderabad, TS, India.

ABSTRACT

Fraud detection is a scenario applicable to many industries such as banking and financial sectors,insurance,healthcare,government agencies and law enforcement and more.There has been a drastic increase in recent years ,pushing fraud detection more important than ever.Hundreds of millions of dollars are lost to fraud every year.Upcoding fraud is one such fraud in which a service provider acquires additional financial gain by coding a service by upgrading it even though the lesser service has been performed.Incorporating artificial intelligence with data mining and statistics help to anticipate and detect these frauds and minimize costs.Using sophisticated data mining tools ,millions of transcations can be searched to spot patterns and detect fraudulent transactions.This paper gives an insight into the various datamining tools which are efficient in detecting upcoding frauds especially in the healthcare insurance sector in India.

I. INTRODUCTION

The term "fraud" implies an intention on the part of some party or an individual presumably planning to commit fraud[1].Different forms of frauds creates threats for the mankind be it social or financial sector .Traditional methods of data analysis was used to detect fraud and it was quite complex and timeconsuming .Fraud comprises of many instances involving repeated contraventions using the same method.Fraud cases can be similar in content and appearance but not identical In banking,fraud may be using stolen Credit cards,falsifying cheques,misleading accounts and more.In insurance,20% to 25% of claims contain some form of fraud,leading to approximately 10% of insurance payout dollars. Fraud can vary from inflated losses to deliberately causing an accident for the payout. With all the different methods of fraud, detecting it

becomes still harder. Data analysis techniques to prevent fraud were first used by the telephone companies,the insurance companies and the banks.Based on a neural network shell ,Falcon fraud assessment system,FICO was successfully implemented in the banking industry.Retail industries also suffer setbacks from frauds. Some supermarkets have started to make use of digitized closed-circuit television (CCTV) together with POS data of most susceptible transactions to fraud.Internet transcations is a big concern nowadays with some research pointing that these trans cation fraud is 12 times higher than in-store fraud. Fraud is an adaptive crime, so it requires special methods of intelligent data analysis to detect and prevent it. These methods exist in the areas of Knowledge Discovery in Databases (KDD), Data Mining, Machine Learning and Statistics[4]. They offer applicable and successful solutions in different areas of fraud crimes.

II. LITERAUTE REVIEW

1.Fraud Detection In Non-Cash Transactions Using Information System Audit (A Case Study Of Government Bank in Bandung Indonesia),Sasongko, Nanang,The Bank toward boosting efficiency within the retail payment system. Bank payment system operators to follow suit in order to encourage public utilization of non-cash payment instrument. In addition bank review implementation of the non-cash payment system. Information systems auditing is the process of collecting and evaluating evidence to determine the safeguards asset of computer system, maintains data integrity, allows organizational goals to be achieved effectively, and uses resources efficiently. This research provides information for fraud-related detection of non-cash transaction, including indications of fraud, internal control analysis, evidence gathering, financial analysis, legal basis for fraud allegations, tracing assets, computer forensic procedures and methods. The sample is one of four government of Bank in Indonesia. Interview is conducted to collect data, applying information systems audit and COBIT level maturity, the result shows that Fraud is manageable, not optimize.

2.Statistical fraud detection: A review, Bolton, Richard J., and David J. Hand, Fraud is increasing dramatically with the expansion of modern technology and the global superhighways of communication, resulting in the loss of billions of dollars worldwide each year. Although prevention technologies are the best way to reduce fraud, fraudsters are adaptive and, given time, will usually find ways to circumvent such

measures. Methodologies for the detection of fraud are essential if we are to catch fraudsters once fraud prevention has failed. Statistics and machine learning provide effective technologies for fraud detection and have been applied successfully to detect activities such as money laundering, e-commerce credit card fraud, telecommunications fraud and computer intrusion, to name but a few. We describe the tools available for statistical fraud detection and the areas in which fraud detection technologies are most used.

3.A survey on the state of healthcare upcoding fraud analysis and detection, Bauder, Richard, Taghi M. Khoshgoftaar, and Naeem Seliya,From its infancy in the 1910s, healthcare group insurance continues to increase, creating a consistently rising burden on the government and taxpayers. The growing number of people enrolled in healthcare programs such as Medicare, along with the enormous volume of money in the healthcare industry, increases the appeal for and risk of fraudulent activities. One such fraud, known as upcoding, is a means by which a provider can obtain additional reimbursement by coding a certain provided service as a more expensive service than what was actually performed. With the proliferation of data mining techniques and the recent and continued availability of public healthcare data, the application of these techniques towards fraud detection, using this increasing cache of data, has the potential to greatly reduce healthcare costs through a more robust detection of upcoding fraud. Presently, there is a sizable body of healthcare fraud detection research available but upcoding fraud studies are limited. Audit data can be difficult to obtain, limiting the usefulness of supervised learning; therefore, other data mining techniques, such as unsupervised learning, must be explored using mostly unlabeled records in order to detect upcoding fraud. This paper is specific to reviewing upcoding fraud analysis and detection research providing an overview of healthcare, upcoding, and a review of the current data mining techniques used therein.

3.Deep learning applications and challenges in big data analytics, Najafabadi, Maryam M., et al, Big Data Analytics and Deep Learning are two high-focus of data science. Big Data has become important as many organizations both public and private have been collecting massive amounts of domain-specific information, which can contain useful information about problems such as national intelligence, cyber security, fraud detection, marketing, and medical informatics. Companies such as Google and

Microsoft are analyzing large volumes of data for business analysis and decisions, impacting existing and future technology. Deep Learning algorithms extract high-level, complex abstractions as data representations through a hierarchical learning process. Complex abstractions are learnt at a given level based on relatively simpler abstractions formulated in the preceding level in the hierarchy. A key benefit of Deep Learning is the analysis and learning of massive amounts of unsupervised data, making it a valuable tool for Big Data Analytics where raw data is largely unlabeled and un-categorized. In the present study, we explore how Deep Learning can be utilized for addressing some important problems in Big Data Analytics, including extracting complex patterns from massive volumes of data, semantic indexing, data tagging, fast information retrieval, and simplifying discriminative tasks. We also investigate some aspects of Deep Learning research that need further exploration to incorporate specific challenges introduced by Big Data Analytics, including streaming data, high-dimensional data, scalability of models, and distributed computing. We conclude by presenting insights into relevant future works by posing some questions, including defining data sampling criteria, domain adaptation modeling, defining criteria for obtaining useful data abstractions, improving semantic indexing, semi-supervised learning, and active learning.

4.A comprehensive survey of data miningbased fraud detection research, Phua, Clifton, Data mining is about finding insights which are statistically reliable, unknown previously, and actionable from data (Elkan, 2001). This data must be available, relevant, adequate, and clean. Also, the data mining problem must be well-defined, cannot be solved by query and reporting tools, and guided by a data mining process model (Lavrac et al, 2004). The term fraud here refers to the abuse of a profit organisation's system without necessarily leading to direct legal consequences. In a competitive environment, fraud can become a business critical problem if it is very prevalent and if the prevention procedures are not fail-safe. Fraud detection, being part of the overall fraud control, automates and helps reduce the manual parts of a screening/checking process. This area has become one of the most established industry/government data mining applications. It is impossible to be absolutely certain about the legitimacy of and intention behind an application or transaction. Given the reality, the best cost effective option is to tease out possible evidences of fraud from the available data using mathematical algorithms. Evolved from numerous research

communities, especially those from developed countries, the analytical engine within these solutions and software are driven by artificial immune systems, artificial intelligence, auditing, database, distributed and parallel computing, econometrics, expert systems, fuzzy logic, genetic algorithms, machine learning, neural networks, pattern recognition, statistics, visualisation and others. There are plenty of specialised fraud detection solutions and software1 which protect businesses such as credit card, e-commerce, insurance, retail, telecommunications industries.

5.Data mining techniques for the detection of fraudulent financial statements, Kirkos, Efstathios, Charalambos Spathis, and Yannis Manolopoulos, This paper explores the effectiveness of Data Mining (DM) classification techniques in detecting firms that issue fraudulent financial statements (FFS) and deals with the identification of factors associated to FFS. In accomplishing the task of management fraud detection, auditors could be facilitated in their work by using Data Mining techniques. This study investigates the usefulness of Decision Trees, Neural Networks and Bayesian Belief Networks in the identification of fraudulent financial statements. The input vector is composed of ratios derived from financial statements. The three models are compared in terms of their performances.

III.EXISTING SYSTEM

Fraud detection is a situation that can occur in a variety of businesses, including the banking and financial sectors, insurance, healthcare, government organizations, and law enforcement. Recent years have seen a sharp increase, making fraud detection more crucial than ever. Every year, fraud causes hundreds of millions of dollars to be lost. Upcoding fraud is one such fraud in which a service provider acquires additional financial gain by coding a service by upgrading it even though the lesser service has been performed

Disadvantages

1. Upcoding is illegal and a fraudulent activity followed by medical providers by cheating the insurance providers and gain more income than they are eligble.
2. This practice is seen to be costly for individual patients as well as for the ins urance payers

IV.PROPOSED SYSTEM

Fraud detection and data mining always goes hand in hand.With the introduction of this technique and with the support of artificial intelligence and machine learning ,fraud

detection has become much simpler and easier and Medical claim fraud detection with the help of data mining plays an important role for an insurance company to detect fraudulent claims Nearly 10% to 20% of the insurance amount is wasted on fraudulent claims.

Advantages

1. with the use of mining data techniques, the wastage can be reduced to a great extent, even though it is difficult to completely eradicate fraud claims.

2. Supervised learning ,Unsupervised and Hybrid learning methods in data mining pave way to efficient methods in detecting faults or anomalies and thus helps in mitigating frauds.

V.MODULES

1. Patient
2. Medical Provider
3. Insurance Provider
4. Bank

MODULE DESCRIPTION

PATIENT

Here patient is a module, patient should register to the application then only he can able to login into the application. After successful login he can perform some operation such as manage his bank account like can send request to bank to take account and can view his status and can deposit money into his account ,
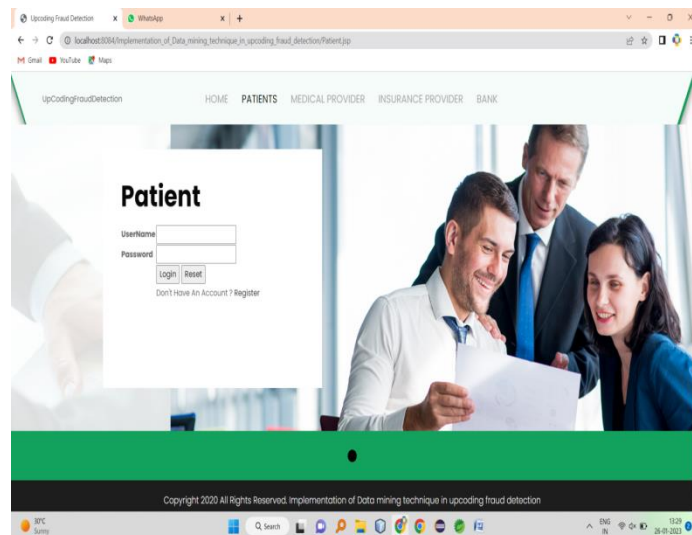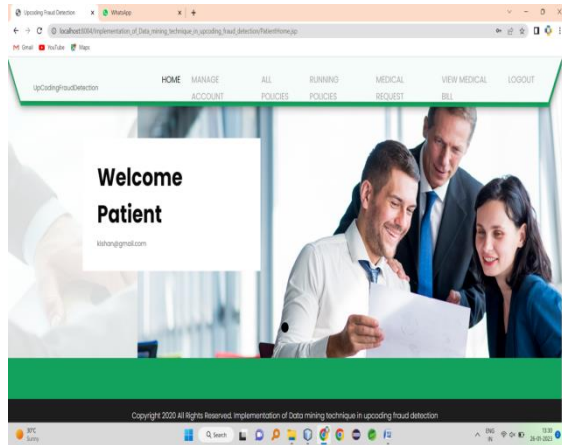
Then the patient can view all policies, can purchase the policy also while purchasing the policy if patient enter more than 3 time wrong cvv then bank have the chance to block his account.
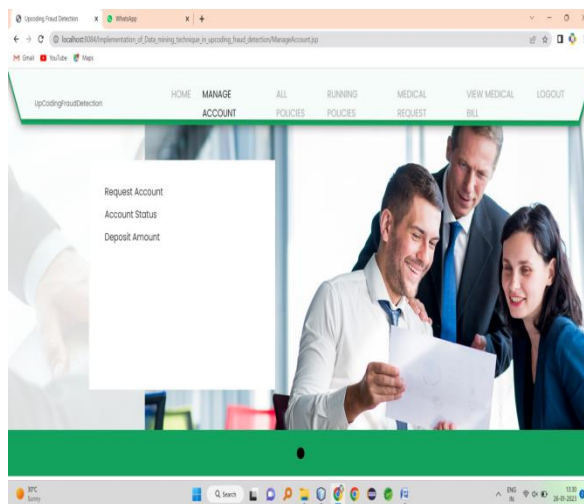
After purchasing the policy patient can send the medical request to the medical provider and generate the treatment bill and can view the medical bill sent by the medical provider and logout.
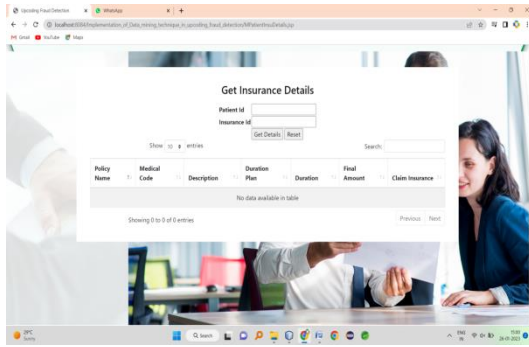
MEDICAL PROVIDER

Here medical provider first as to take one account with specified specialization then login into his account, after successful login he can view medical request, can generate the medical bill and send the insurance details request to patient and get the patient insurance details and apply insurance behalf of patient and get the claim amount if everything is clear then at last logout.
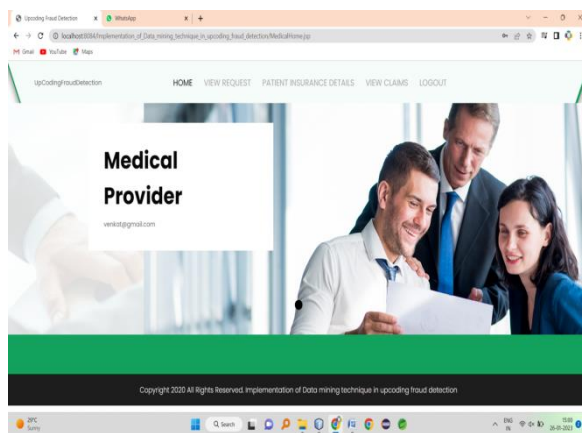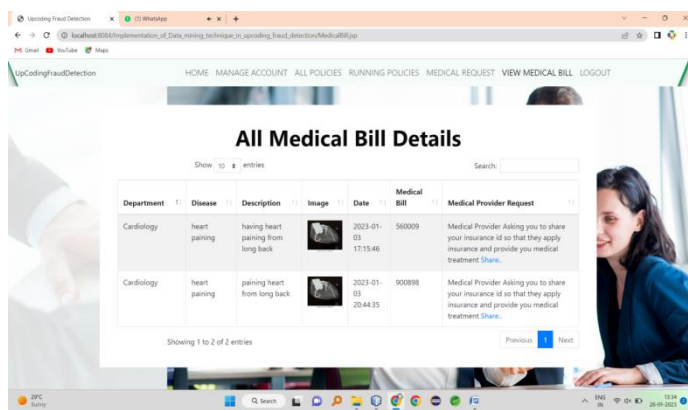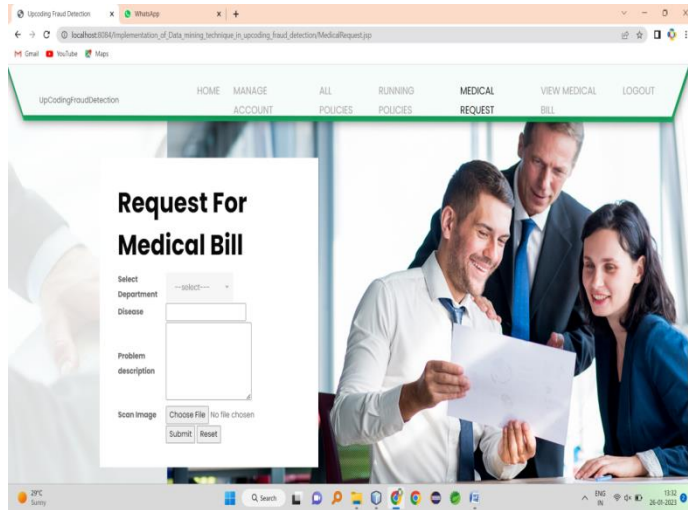
INSURANCE PROVIDER

Here insurance provider also take one account into this application after successful login he can perform some operations like he can add policies and can view policies and can view all policy holder.

And the insurance provider can verify the insurance claim request, if the insurance amount is less than medical bill using data mining filtering technique our system will identify and show to the insurance provider, here insurance provider can block those claim request and logout
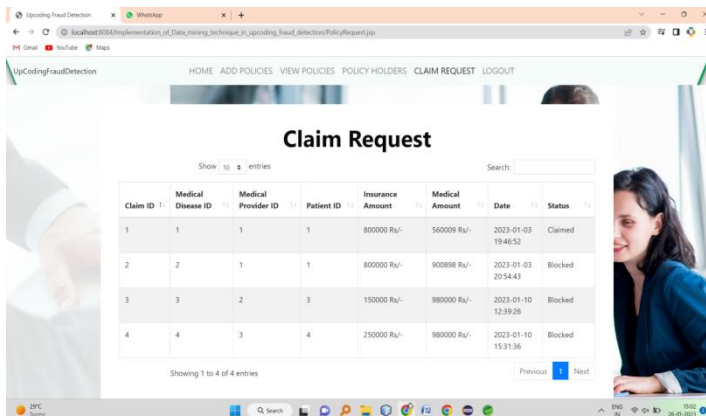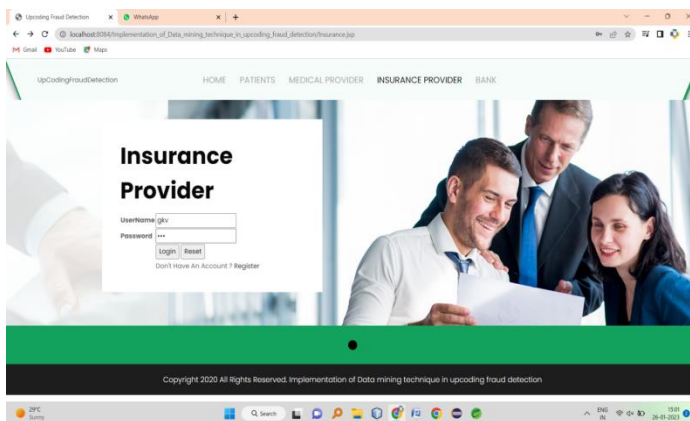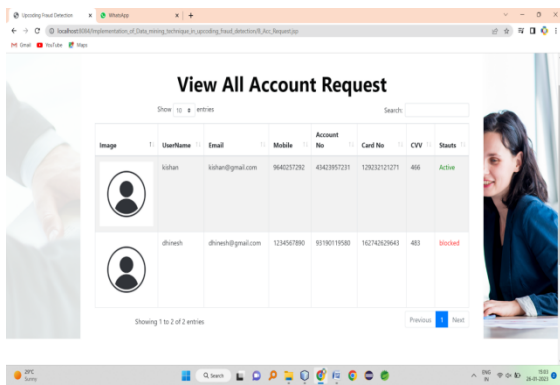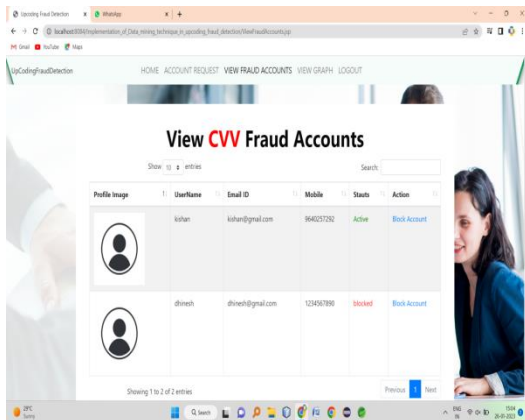

BANK

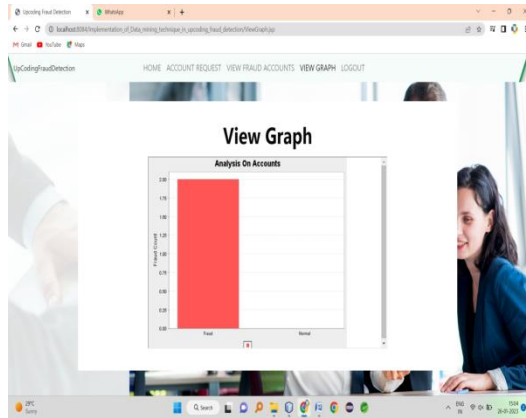Here bank is a module, bank can directly login with specified details and after successful login
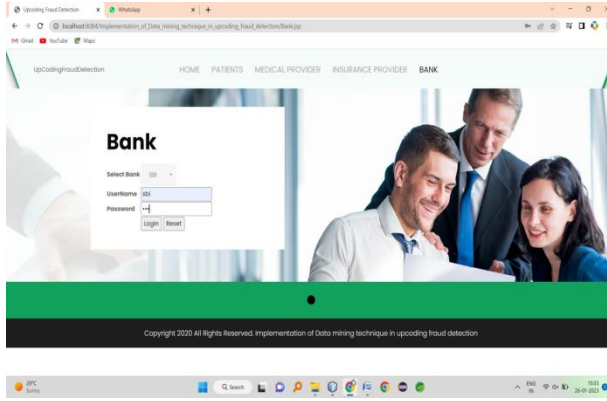
he can perform some operation such as view account request,

view fraud accounts, view graph and

logout

VI.CONCLUSION

Fraud including upcoding puts a heavy financial burden not only to the insurer but also to the customer as this would increase the premium rates payable to the insurer.So efficient mechanisms has to be formulated and implemented to detect these financial frauds as well as to mitigate it.

Healthcare fraud detection studies are limited using supervised and unsupervised learning methodologies and is very few in the case of upcoding healthcare fraud detection.Linear regression ,mixed logit,Bayesian models are the supervised techniques used for upcoding fraud detection .A combination of subgroup creation via decision tree and Fisher's Exact Test are done using the unsupervised learning techniques.The application of additional learning and classification technique in this field will surely pave way for more innovative research options to reduce the

risk of upcoding frauds.

VII.REFERENCES

[1] Sasongko, Nanang. "Fraud Detection In Non-Cash Transactions Using Information System Audit (A Case Study Of Government Bank in Bandung Indonesia)." ASEAN/Asian Academic Society International Conference Proceeding Series. 2015.

[2] Palshikar, Girish Keshav. "The hidden truth-frauds and their control: A critical application for business intelligence." Intelligent Enterprise 5.9 (2002): 46-51.

[3]https://en.wikipedia.org/wiki/Data_analysis_techniques_for_fra ud_detection

[4] Bolton, Richard J., and David J. Hand. "Statistical fraud detection: A review." Statistical science (2002): 235-249.

JOURNAL OF CURRENT SCIENCE

[5] Bauder, Richard, Taghi M. Khoshgoftaar, and Naeem Seliya. "A survey on the state of healthcare upcoding fraud analysis and detection."Health Services and Outcomes Research Methodology: 1-25.

[6] https://en.wikipedia.org/wiki/ICD-10

[7] http://www.who.int/classifications/icd/en/

[8] Najafabadi, Maryam M., et al. "Deep learning applications and challenges in big data analytics." Journal of Big Data 2.1 (2015): 1.

[9] Phua, Clifton, et al. "A comprehensive survey of data miningbased fraud detection research." arXiv preprint arXiv:1009.6119 (2010).

[10] Kirkos, Efstathios, Charalambos Spathis, and Yannis Manolopoulos. "Data mining techniques for the detection of fraudulent financial statements." Expert systems with applications 32.4 (2007): 995-1003.