INCORPORATING DATA MINING METHODS FOR ENHANCINGFRAUD DETECTIONWITHIN FINANCIAL SECTORS

[1] Dr.T.Charan Singh, [2] M.Tharun Kumar, [3] P.Karthik Goud, [4] N.Prathik Reddy, [5] Md.Abdul Mudassir
[1] Professor,[2345]B.Tech Students
Department Of Computer Science & Engineering
Sri Indu College Of Engineering & Technology,Sheriguda, Ibrahimpatnam

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 transactions can be searched to spot patterns and detect fraudulent transactions. This paper gives an insight into the various data mining 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. 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 time consuming. 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 transactions 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. They offer applicable and successful solutions in different areas of fraud crimes Data Mining is associated with supervised learning based on training data of known fraud and legitimate cases and unsupervised learning with data that are not labeled to be fraud or legitimate. Bedford’s law can be interpreted as an example of unsupervised learning (Bolton et al. 2002). The direct application of these methods to forensic accounting is limited due to almost complete nonexistence of large sets of fraud training data (Bolton et al. 2002; Jensen, 1997). Insurance fraud, credit card fraud, telecommunications fraud.

Insurance fraud is common in automobile, travel. The Uniform Suspected Insurance Fraud Reporting Form, adopted by the NAIC Antifraud Task Force 2003, replaced the prior Task Force form. This form standardizes insurance fraud data for the insurance industry and makes it easier to report and track. Fraud detection involves three types of offenders (Baldock, 1997):

i) Criminal offenders.

ii) Organized criminal offenders who are responsible for major fraud.

iii) Offenders who commit fraud.

II. LITERATURE SURVEY

1. Title: DATA MINING TECHNIQUES IN FRAUD DETECTION

Author: Rekha Bhowmik.

Abstract: The paper presents application of data mining techniques to fraud analysis. We present some classification and prediction data mining techniques which we consider important to handle fraud detection. There exist a number of data mining algorithms and we present statistics- based algorithm, decision tree based algorithm and rule-based algorithm. We present Bayesian classification model to detect fraud in automobile insurance. Naïve Bayesian visualization is selected to analyze and interpret the classifier predictions. We illustrate how ROC curves can be deployed for model assessment in order to provide a more intuitive analysis of the models.

2. Title: A COMPLETE LITERATURE REVIEW ON FINANCIAL FRAUD DETECTION APPLYING DATA MINING TECHNIQUES

Author: Subhas Barman∗,1 Utpal Pal Md. Asif Sarfaraj Biswajit Biswas Animesh Mahata Palash Mandal

Abstract: Financial fraud is defined as unlawful or criminal duplicity attempted to result organizational or personal gain. This is a terrible threat to the economics of a firm, corporate sector, Government or ordinary customers. Several processes exist to detect different types of financial fraud. But, due to inefficiency of those processes, researchers leverage data mining techniques to detect financial fraud.

In the practice of data mining, six data mining classes (Classification, Clustering, Visualization, Prediction, Regression, Outlier detection) have been used as a core or base while different techniques (K-nearest neighbour, Decision tree, Fuzzy logic, logistic model, Bayesian belief network, Naïve Bayes, Beneish M-Score model, Benford's law, Altman Z-score) have been applied to improve accuracy of fraud detection. In this paper, existing financial fraud detection techniques are compared with the advantage and limitations of the techniques.

3. Title: A COMPREHENSIVE SURVEY OF DATA MINING-BASED FRAUD DETECTION RESEARCH

Author: Clifton phua, vincent lee , kate smith1 & ross gayler School of Business Systems, Faculty of Information Technology, Monash University

Abstract: This survey paper categorises, compares, and summarises from almost all published technical and review articles in automated fraud detection within the last 10 years. It defines the professional fraudster, formalises the main types and subtypes of known fraud, and presents the nature of data evidence collected within affected industries.

Within the business context of mining the data to achieve higher cost savings, this research presents methods and techniques together with their problems. Compared to all related reviews on fraud detection, this survey covers much more technical articles and is the only one, to the best of our knowledge, which proposes alternative data and solutions from related domains.

4. Title: FRAUD DETECTION USING THE FRAUD TRIANGLE THEORY AND DATA MINING TECHNIQUES: A LITERATURE REVIEW

Author: Marco Sánchez-Aguayo , Luis Urquiza-Aguiar , and José Estrada-Jiménez

Abstract: Fraud entails deception in order to obtain illegal gains; thus, it is mainly evidenced within financial institutions and is a matter of general interest. The problem is particularly complex, since perpetrators of fraud could belong to any position, from top managers to payroll employees. Fraud detection has traditionally been performed by auditors, who mainly employ manual techniques. These could take too long to process fraud-related evidence.We used the Kitchenham methodology to analyze the research works related to fraud detection from the last decade. This review provides evidence that fraud is an area of active investigation. Several works related to fraud detection using machine learning techniques were identified without the evidence that they incorporated the fraud triangle as a method for more efficient analysis.

III. SYSTEM ANALYSIS & DESIGN

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. Existing System uses Data Analysis methods to detect fraud detection in financial sectors but, the problem arises in the analyzing of large data of big corporate companies.

Frauds are becoming very must easy to do in any of the sectors mainly in financial sectors insurance frauds or increasing everyday. Internet transactions is a big concern nowadays with some research pointing that these transcation fraud is 12 times higher than in-store fraud. Fraudulent insurance claims are one of the biggest preventable losses that hurt insurers world wide.The P&C segment accounts for the most fraudulent insurance claims, with auto insurance and workers' compensation making up the biggest percentage of fraudulent claims that annually impact the insurance business

DISADVANTAGES

- Upcoding is illegal and a fraudulent activity followed by medical providers by cheating the insurance providers and gain more income than they are eligible.
- This practice is seen to be costly for individual patients as well as for the insurance pay3rs
- Fraud can distort markets by giving fraudsters a competitive advantage and driving out legitimate businesses.
- Fraud can cause immediate and long-term environmental damage by polluting the environment and destroying ecosystems and biodiversity. It can also lead to significant clean-up expenses

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

- 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.
- 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.
- Data mining techniques were used to detect fraud across different financial applications such as health insurance and credit card.
- Data mining with its wide variety of techniques is able to juice out a lot of useful information from a large set of data.
- With its ability to find useful knowledge from a given data, it is a potent technique to identify abnormal patterns in data and any underlying unwanted activity.
- Industries like Insurance, Banking, Credit Card, and Telecom are most vulnerable to financial fraud with large sets of data.

SYSTEM ARCHITECTURE

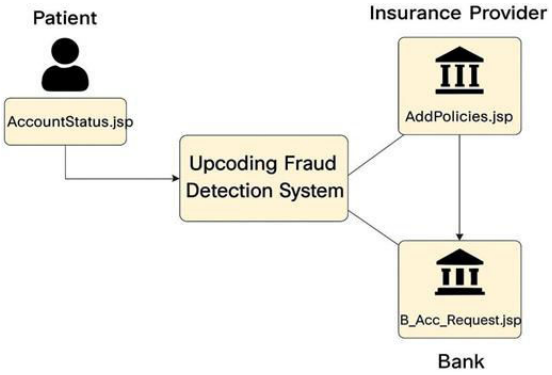

Fig: System Architecture

IV. IMPLEMENTATION

MODULES

- Patient
- Medical Provider
- Insurance Provider
- 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.

Patient can perform these following functions in patient module:

- Register
- Login
- Manage account
- All policies
- Running policies
- medical request
- View medical bill
- 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.

Medical provider can perform these following functions in medical provider module:

- Register
- Login
- View request
- Patient insurance details
- View claims
- 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

Insurance provider can perform these following functions in insurance provider module:

- Register
- Login
- Add policies
- View policies
- Policy holder
- Claim request
- 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 Bank make the main role in detecting fraud ,any fraudulent activities are only detected by bank module.only bank can block or activate any account

Bank can perform these following functions in bank module:

- Login
- Account request
- View fraud account
- View graph
- logout

V. SCREENSHOTS

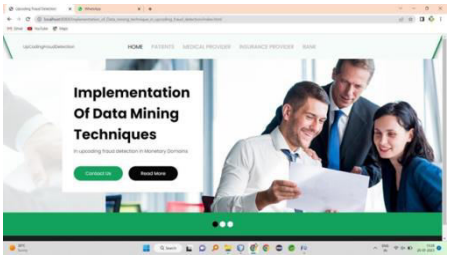

Fig 1 : Index Page

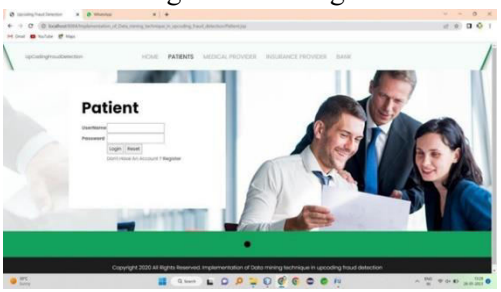

Fig : Patient Login

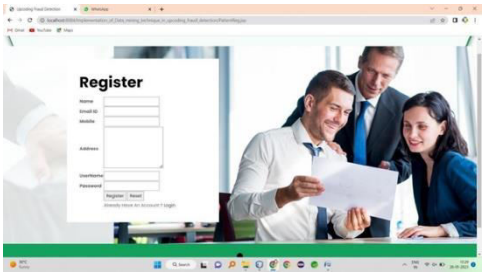

Fig : Patient Register

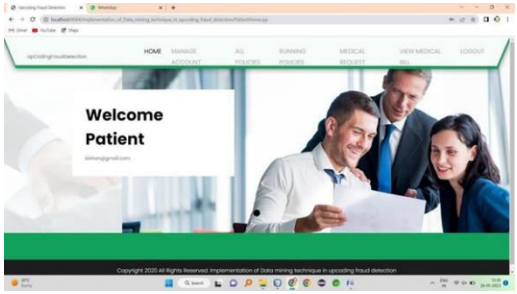

Fig: Patient home page

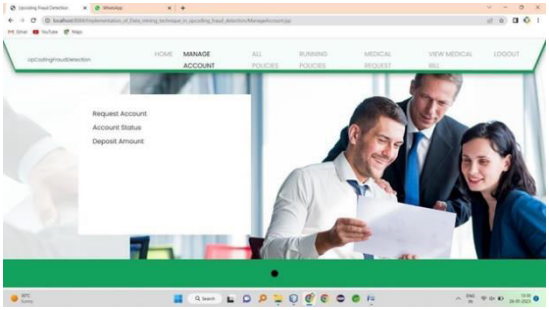

Fig :Manage Account

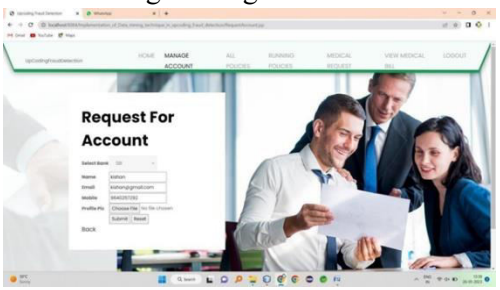

Fig : Request for Account

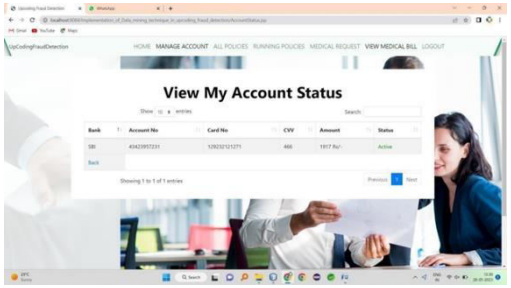

Fig: Account status

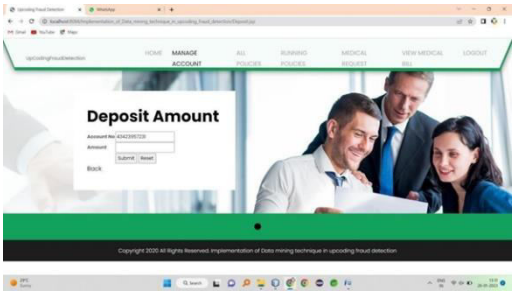

Fig : Deposit amount

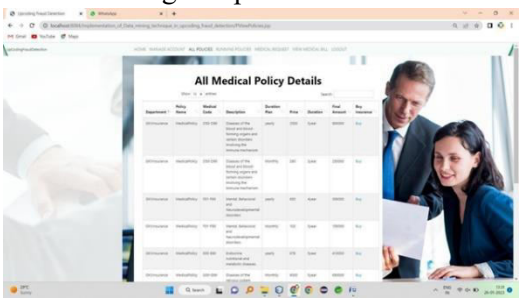

Fig : All medical policies

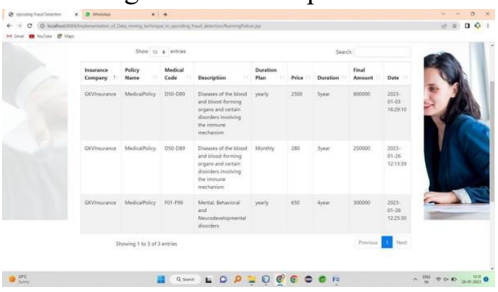

Fig : All running policies

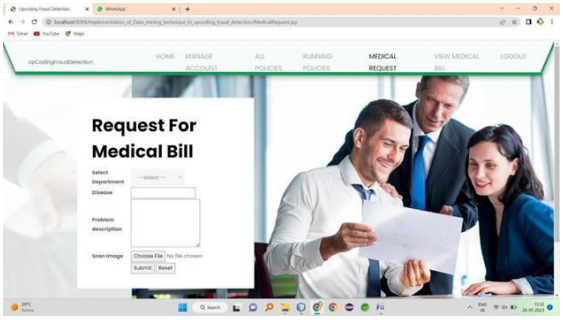

Fig: Medical request

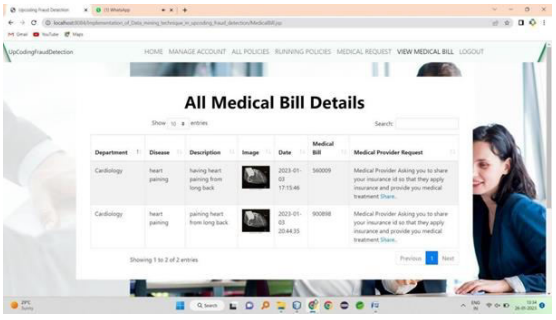

Fig: Medical bill

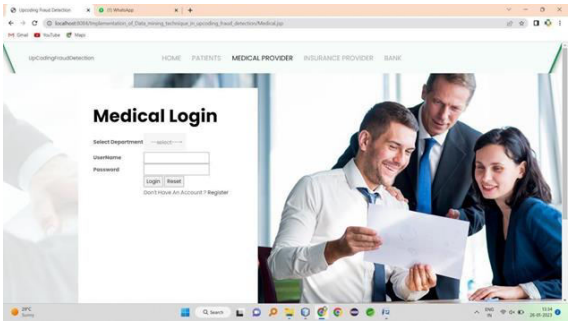

Fig: Medical login

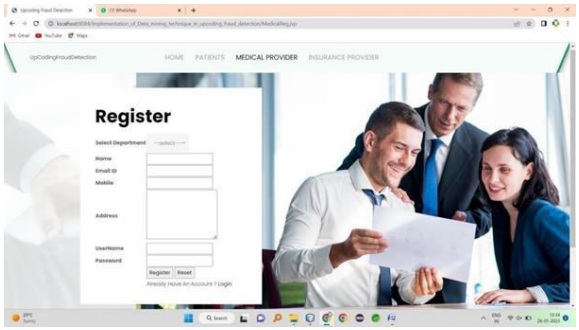

Fig : Medical provider register

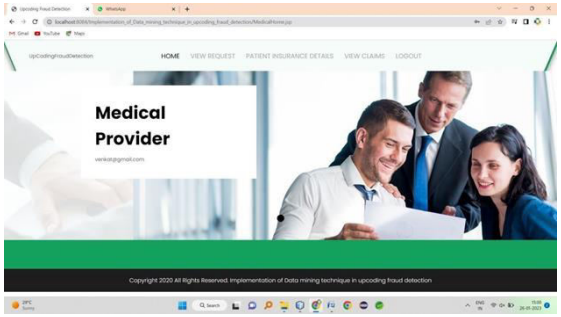

Fig : Medical provider home page

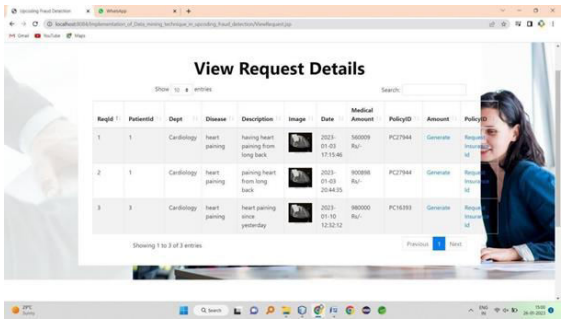

Fig: View request from patient

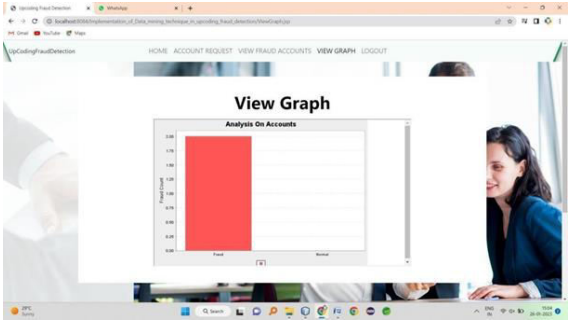

Fig:View fraud accounts in graph

VI. CONCLUSION

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 log it,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.

FUTURE SCOPE

The future scope for using data mining methods to detect fraud in the financial sector is promising, as technological advancements, evolving fraud techniques, and increased data availability continue to drive innovation in this field. Here are some key areas of future development and opportunities for data mining in fraud detection within the financial sector:

Machine Learning and AI Integration:The integration of machine learning and artificial intelligence (AI) into data mining techniques will enhance fraud detection accuracy. Deep learning models, natural language processing (NLP), and anomaly detection algorithms will play a more significant role in identifying complex fraud patterns.

Big Data and Real-time Analysis: As financial data volumes continue to grow, data mining will need to adapt to handle big data. Real-time analysis of financial transactions and customer behavior will become increasingly important for immediate fraud detection and prevention.

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. https://en.wikipedia.org/wiki/Data_analysis_techniques_for_fra ud_detection
3. Bolton, Richard J., and David J. Hand. "Statistical fraud detection: A review." Statistical science (2002): 235-249.
4. 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.
5. Najafabadi, Maryam M., et al. "Deep learning applications and challenges in big data analytics." Journal of Big Data 2.1 (2015)
6. Phua, Clifton, et al. "A comprehensive survey of data mining based fraud detection research." arXiv preprint arXiv:1009.6119 (2010).
7. 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