PRUDENT FRAUD DETECTION USING MACHINE LEARNING

K.SUPARNA [1], VENDRA JAGADEESH[2]

[1] **Assistant Professor MSc (CS), DEPT,** Dantuluri Narayana Raju College, **Bhimavaram, Andhra Pradesh**
Email id:- suparnakalidindi@gmail.com
[2]**PG Student of MSc Computer Science,** Dantuluri Narayana Raju College,
Bhimavaram, Andhra Pradesh
Email id :- vendrajagadeesh327@gmail.com

ABSTRACT

Most commercial Fraud Detection components of Internet banking systems use some kind of hybrid setup usually comprising a Rule-Base and an Artificial Neural Network. Such rule bases have been criticised for a lack of innovation in their approach to Knowledge Acquisition and maintenance. Furthermore, the systems are brittle; they have no way of knowing when a previously unseen set of fraud patterns is beyond their current knowledge. This limitation may have far reaching consequences in an online banking system. This paper presents a viable alternative to brittleness in Knowledge Based Systems; a potential milestone in the rapid detection of unique and novel fraud patterns in Internet banking. The experiments conducted with real online banking transaction log files suggest that Prudent based fraud detection may be a worthy alternative in online banking. This limitation may have far reaching consequences in an online banking system. This paper presents a viable alternative to brittleness in Knowledge Based Systems; a potential milestone in the rapid detection of unique and novel fraud patterns in Internet banking. The experiments conducted with real online banking transaction log files suggest that Prudent based fraud detection may be a worthy alternative in online banking. transactional data in real-time. By amalgamating both supervised and unsupervised learning methods, the proposed system strives to not only identify fraudulent activities but also to mitigate the occurrence of false positives, thus elevating the overall security and efficiency of finass various phases, commencing with meticulous data preprocessing to eliminate noise, rectify outliers, standardize features, and transform categorical data into a format conducive to machine learning analysis. Subsequently, the system employs advanced feature engineering techniques to extract pertinent information from the dataset, resulting in enriched features that enhance

Most commercial Fraud Detection components of Internet banking systems use some kind of hybrid setup usually comprising a Rule-Base and an Artificial Neural Network. Such rule bases have been criticised for a lack of innovation in their approach to Knowledge Acquisition and maintenance. Furthermore, the systems are brittle; they have no way of knowing when a previously unseen set of fraud patterns is beyond their current knowledge. This limitation may have far reaching consequences in an online banking system. This paper presents a viable alternative to brittleness in Knowledge Based Systems; a potential milestone in the rapid detection of unique and novel fraud patterns in Internet banking. The experiments conducted with real online banking transaction log files suggest that Prudent based fraud detection may be a worthy alternative in online banking. This limitation may have far reaching consequences in an online banking system. This paper presents a viable alternative to brittleness in Knowledge Based Systems; a potential milestone in the rapid detection of unique and novel fraud

patterns in Internet banking. The experiments conducted with real online banking transaction log files suggest that Prudent based fraud detection may be a worthy alternative in online banking. transactional data in real-time. By amalgamating both supervised and unsupervised learning methods, the proposed system strives to not only identify fraudulent activities but also to mitigate the occurrence of false positives, thus elevating the overall security and efficiency of finass various phases, commencing with meticulous data preprocessing to eliminate noise, rectify outliers, standardize features, and transform categorical data into a format conducive to machine learning analysis. Subsequently, the system employs advanced feature engineering techniques to extract pertinent information from the dataset, resulting in enriched features that enhance

1 INTRODUCTION

According to The American Heritage dictionary, second college edition, fraud is defined as a deception deliberately practiced to secure unfair unlawful gain. Fraud detection is the recognition of symptoms of fraud where no prior suspicion or tendency to fraud exists. Examples include insurance fraud, credit card fraud and accounting fraud. Data from the Nigeria Inter-Bank Settlement System (NIBSS) has revealed that fraudulent transactions in the banking sector at its peak. Fraud has evolved from being committed by casual fraudsters to being committed by organized crime and fraud rings that use sophisticated methods to take over control of accounts and commit fraud. Some 6.8 million Americans were victimized by card fraud in 2007, according to Javelin research. Such fraud on existing accounts accounted for more than $3 billion in losses in 2007. The Nilson Report estimates the cost to the industry to be $4.84 billion. Javelin estimates the losses at more than six times that amount – some $30.6 billion in 2007. Of course, fraud is not a domestic product as it's everywhere. For instance, card fraud losses cost UK economy GBP 423 million in 2006. Credit card fraud accounts for the biggest cut of the $600 million that airlines lose each year globally.

OVERVIEW

Fraud detection is a set of activities undertaken to prevent money or property from being obtained through false pretenses. Fraud detection is applied to many industries such as banking or insurance. In banking, fraud may include forging checks or using stolen credit cards.

With an unlimited and rising number of ways someone can commit fraud, detection can be difficult. Activities such as reorganization, downsizing, moving to new information systems or encountering a cyber security breach could weaken an organization's ability to detect fraud. Techniques such as real-time monitoring for fraud are recommended. Organizations should look for fraud in financial transactions, locations, devices used, initiated sessions and authentication systems.

Fraud can be committed in different ways and different settings. For example, fraud can be committed in banking, insurance, government and healthcare sectors. A common type of banking fraud is customer account takeover. This is when someone illegally gains access to a victim's bank account using bots. Other examples of fraud in banking include the use of malicious applications, the use of false identities, money laundering, credit card fraud and mobile fraud.

Government fraud is committing fraud against federal agencies such as the

U.S. Department of Health and Human Services, Department of Transportation, Department of Education or Department of Energy. Types of government fraud include billing for unnecessary procedures, overcharging for items that cost less, providing old equipment when billing for new equipment and reporting hours worked for a worker that does not exist.

2. LITERATURE SURVEY AND RELATED WORK

Fraud detection has been usually seen as a data mining problem where the objective is to correctly classify the transactions as legitimate or fraudulent. For classification problems many performance measures are defined most of which are related with correct number of cases classified correctly.

A more appropriate measure is needed due to the inherent structure of credit card transactions. When a card is copied or stolen or lost and captured by fraudsters it is usually used until its available limit is depleted. Thus, rather than the number of correctly classified transactions, a solution which minimizes the total available limit on cards subject to fraud is more prominent.Since the fraud detection problem has mostly been defined as a classification problem, in addition to some statistical approaches many data mining algorithms have

been proposed to solve it. Among these, decision trees and artificial neural networks are the most popular ones. The study of Bolton and Hand provides a good summary of literature on fraud detection problems.

However, when the problem is approached as a classification problem with variable misclassification costs as discussed above, the classical data mining algorithms are not directly applicable; either some modifications should be made on them or new algorithms developed specifically for this purpose are needed. An alternative approach could be trying to make use of general purpose meta heuristic approaches like genetic algorithms.

2.1 Neural nets versus conventional techniques in credit scoring in Egyptian banking

The number of Non-Performing Loans has increased in recent years, paralleling the current financial crisis, thus increasing the importance of credit scoring models. This study proposes a three stage hybrid Adaptive Neuro Fuzzy Inference System credit scoring model, which is based on statistical techniques and Neuro Fuzzy. The proposed model's performance was compared with conventional and commonly utilized models. The credit scoring models are tested using a 10-fold cross-validation process with the credit card data of an international bank operating in Turkey. Results demonstrate that the proposed model consistently performs better than the Linear Discriminant Analysis, Logistic Regression Analysis, and Artificial Neural Network (ANN) approaches, in terms of average correct classification rate and estimated misclassification cost. As with ANN, the proposed model has learning ability; unlike ANN, the model does not stay in a black box. In the proposed model, the interpretation of independent variables may provide valuable information for bankers and consumers, especially in the explanation of why credit applications are rejected.

2.2 A credit scoring model for Vietnams retail banking market

As banking markets in developing countries are maturing, banks face competition not only from other domestic banks but also from sophisticated foreign banks. Given the substantial growth of consumer credit and increased

regulatory attention to risk management, the development of a well-functioning credit assessment framework is essential. As part of such a framework, we propose a credit scoring model for Vietnamese retail loans. First, we show how to identify those borrower characteristics that should be part of a credit scoring model. Second, we illustrate how such a model can be calibrated to achieve the strategic objectives of the bank.

2.3 Statistical classification methods in consumer credit scoring

Credit scoring is the term used to describe formal statistical methods used for classifying applicants for credit into _good' and _bad' risk classes. Such methods have become increasingly important with the dramatic growth in consumer credit in recent years. A wide range of statistical methods has been applied, though the literature available to the public is limited for reasons of commercial confidentiality. Particular problems arising in the credit scoring context are examined and the statistical methods which have been applied are reviewed.

2.4 A comparison of neural networks and linear scoring models in the credit union environment

The purpose of the present paper is to explore the ability of neural networks such as multilayer perceptron 's and modular neural networks, and traditional techniques such as linear discriminant analysis and logistic regression, in building credit scoring models in the credit union environment. Also, since funding and small sample size often preclude the use of customized credit scoring models at small credit unions, we investigate the performance of generic models and compare them with customized models. Our results indicate that customized neural networks offer a very promising avenue if the measure of performance is percentage of badloans correctly classified. However, if the measure of performance is percentage of good and bad loans correctly classified, logistic regression models are comparable to the neural networks approach. The performance of generic models was not as good as the customized models, particularly when it came to correctly classifying bad loans. Although we found significant differences in the results for the three credit unions, our modular neural network could not accommodate these differences, indicating that more innovative architectures might be necessary for building effective generic models.

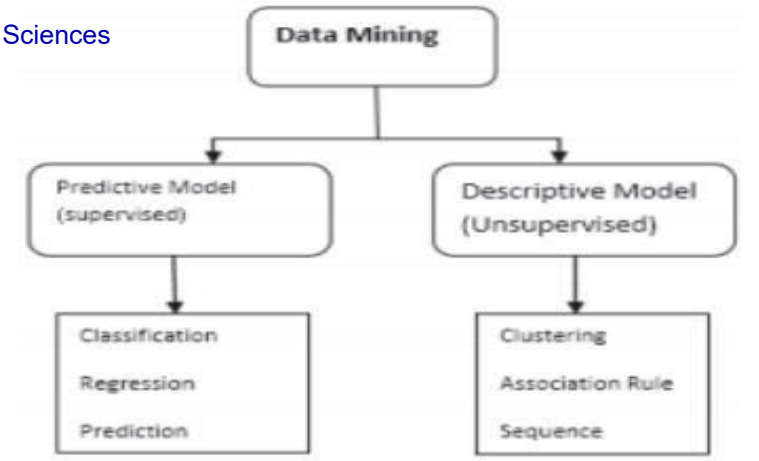
2.5 Credit Scoring Methods. Czech Journal of Economics and Finance

The paper reviews the best-developed and most frequently applied methods of credit scoring employed by commercial banks when evaluating loan applications. The authors concentrate on retail loans – applied research in this segment is limited, though there has been a sharp increase in the volume of loans to retail clients in recent years. Logit analysis is identified as the most frequent credit- scoring method used by banks. However, other nonparametric methods are widespread in terms of pattern recognition. The methods reviewed have potential for application in post-transition countries.

2.6 A survey of credit and behavioural scoring: forecasting: financial risk of lending to customers

Credit scoring and behavioural scoring are the techniques that help organisations decide whether to grant credit to consumers who apply to them. This article surveys the techniques used — both statistical and operational research based — to support these decisions. It also discusses the need to incorporate economic conditions into the scoring systems and the way the systems could change from estimating the probability of a consumer defaulting to estimating the profit a consumer will bring to the lending organisation — two of the major developments being attempted in the area. It points out how successful has been this under-rights reserved.

3. PROPOSED WORK AND ALGORITHM



In proposed methodology, Detection of fraudulent activity is thus critical to control these costs. This paper hereby addresses bank fraud detection via the use of machine learning techniques; association, clustering, forecasting, and classification to analyze the customer data to identify the patterns that can lead to frauds. Upon identification of the patterns, adding a higher level of verification/authentication to banking processes can be added. These kinds of frauds can be credit card fraud, insurance fraud, accounting fraud, etc. which may lead to the financial loss to the bank or the customers. Thus, detection of these kinds of frauds are very important. Fraud detection in banking sector is based on the machine learning techniques and their collective analysis from the past experiences and the probability of how the fraudsters can steal from customers and banks. Therefore, this paper addresses the analysis of data mining techniques of how to detect frauds and overcoming it in banking sector.

ADVANTAGES OF PROPOSED SYSTEM

To eliminate real time fraud to the lowest level.

To increase the confidence of customers in the banking system especiallyfor online transactions.

To discourage fraudsters (both present and intending ones)

4. METHODOLOGIES

MACHINELEARNING

Machine learning could be a subfield of computer science (AI). The goal of machine learning typically is to know the structure information of knowledge of information and match that data into models which will be understood and used by folks. Although machine learning could be a field inside technology, it differs from ancient processapproaches.

In ancient computing, algorithms are sets of expressly programmed directions employed by computers to calculate or downside solve. Machine learning algorithms instead give computers to coach on knowledge inputs and use applied math analysis so as to output values that fall inside a particular vary. thanks to this, machine

learning facilitates computers in building models from sample knowledge tomodify decision-making processes supported knowledge inputs.

MACHINE LEARNING STRATEGIES

In machine learning, tasks square measure typically classified into broad classes. These classes square measure supported however learning is received or however feedback on the educational is given to the system developed. Two of the foremost wide adopted machine learning strategies square measure supervised learning that trains algorithms supported example input and output information that's tagged by humans, and unattended learning that provides the algorithmic program with no tagged information so as to permit it to search out structure at intervals its computer file.

SUPERVISED LEARNING

In supervised learning, the pc is given example inputs that square measure labelled with their desired outputs. The aim of this technique is for the algorithmic program to be ready to ―learn‖ by comparison its actual output with the ―taught‖ outputs to search out errors, and modify the model consequently. Supervised learning thus uses patterns to predict label values on extra unlabeled information. For example, with supervised learning, an algorithm may be fed data with images of sharks labelled as fish and images of oceans labelled as water. By being trained on this data, the supervised learning algorithm should be able to later identify unlabeled shark images as fish and unlabeled ocean images as water.

A common use case of supervised learning is to use historical information to predict statistically probably future events. It's going to use historical stock exchange info to anticipate approaching fluctuations or be used to filter spam emails. In supervised learning, labeled photos of dogs are often used as input file to classify unlabeled photos of dogs.

UNATTENDED LEARNING

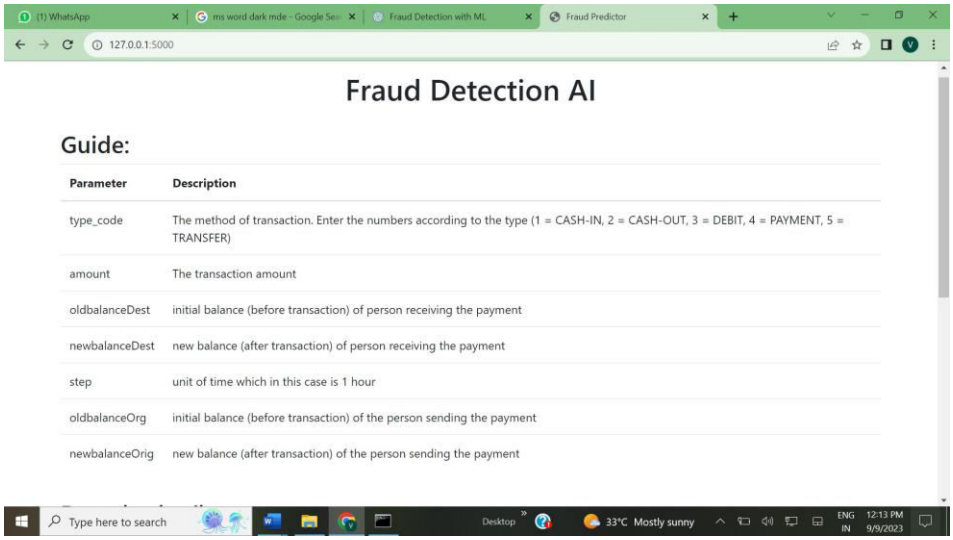
In unattended learning, information is unlabeled, that the learning rule is left to seek out commonalities among its input file. The goal of unattended learning is also as easy as discovering hidden patterns at intervals a dataset;however it should even have a goal of feature learning, that permits the procedure machine to mechanically discover the representations that square measure required to classify data.

Unsupervised learning is usually used for transactional information. You will have an oversized dataset of consumers and their purchases, however as a person's you'll probably not be able to add up of what similar attributes will be drawn from client profiles and their styles of purchases.

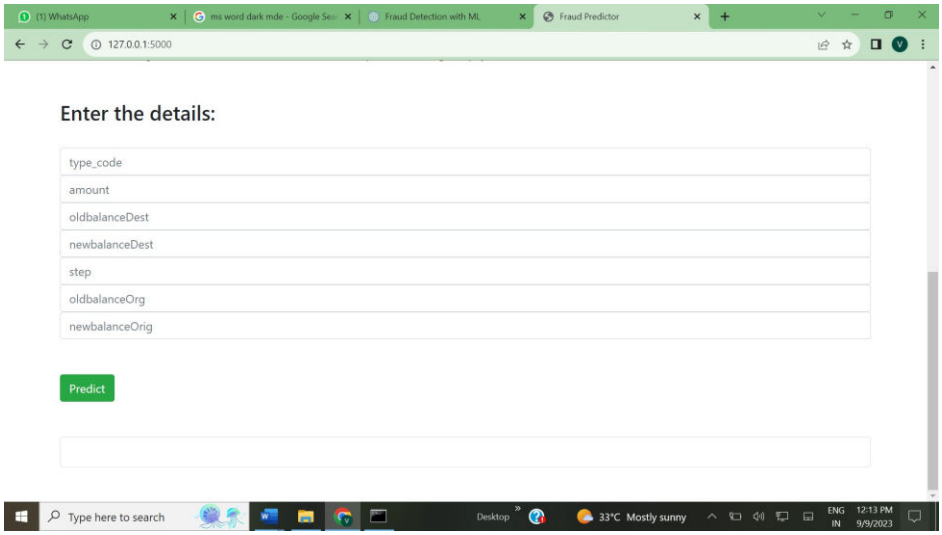
With this information fed into Associate in Nursing unattended learning rule, it should be determined that ladies of a definite age vary UN agency obtain unscented soaps square measure probably to be pregnant, and so a promoting campaignassociated with physiological condition and baby will be merchandised.

5.RESULTS AND DISCUSSION SCREENSHOTS

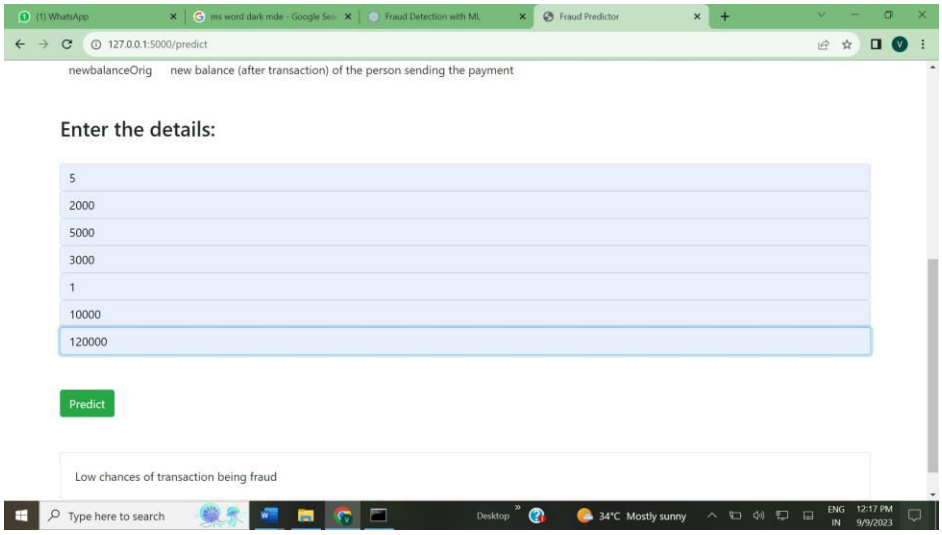
5.1 HOME SCREEN



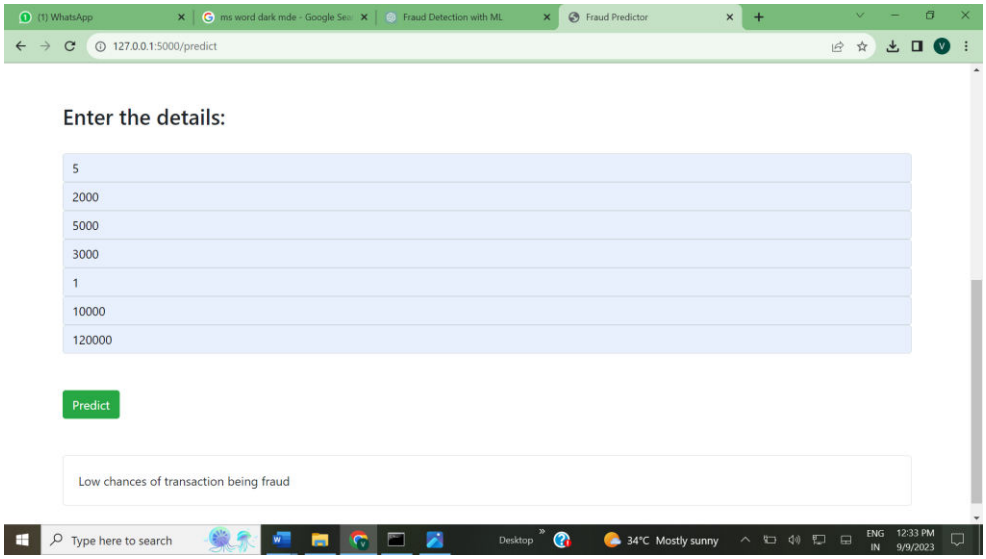
5.2 ENTER DETAILS



5.3 UPLOAD DETAILS



5.4 **OUTPUT**



6.CONCLUSION

Machine Learning is a technique used to extract vital information from existing huge amount of data and enable better decision-making for the banking and retail industries. They use data warehousing to combine various data from databases into an acceptable format so that the data can be mined. The data is then analyzed and the information that is captured is used throughout the organization to support decision-making. Data Mining techniques are very useful to the banking sector for better targeting and acquiring new customers, most valuable customer retention, automatic credit approval which is used for fraud prevention, fraud detection in real time, providing segment based products, analysis of the customers, transaction patterns over time for better retention and relationship, risk management and marketing.Advanced Machine Learning Models: In the ever-evolving landscape of financial fraud, the application of advanced machine learning models is a promising avenue. Future developments in this project should explore the use of more sophisticated models, such as deep neural networks,

recurrent neural networks (RNNs), or convolutional neural networks (CNNs). These models have the potential to capture intricate patterns and anomalies in transaction data, thereby accuracy

7.REFERENCES AND BIBILOGRAPHY

1. S. M. Darwish, "An intelligent credit card fraud detection approach based on semantic fusion of two classifiers," (in English), Soft Computing, Article vol. 24,no. 2, pp. 1243-1253, Jan 2020.
2. A. Eshghi and M. Kargari, "Introducing a new method for the fusion of fraud evidence in banking transactions with regards to uncertainty," (in English), Expert Systems with Applications, Article vol. 121, pp. 382-392, May 2019.
3. S. Hossain, A. Abtahee, I. Kashem, M. M. Hoque, and I. H. Sarker, "Crime Prediction Using Spatio-Temporal Data," in Computing Science, Communication and Security, Singapore, 2020, pp. 277-289: Springer Singapore.
4. M. Zamini and S. M. H. Hasheminejad, "A comprehensive survey of anomaly detection in banking, wireless sensor networks, social networks, and healthcare," (in English), Intelligent Decision Technologies-Netherlands, Article vol. 13, no. 2, pp. 229-270, 2019.
5. I. Gonzalez-Carrasco, J. L. Jimenez-Marquez, J. L. Lopez-Cuadrado, and B. Ruiz-Mezcua, "Automatic detection of relationships between banking operations using machine learning," (in English), Information Sciences, Article vol. 485, pp. 319-346, Jun 2019.
6. M. Pohoretskyi, D. Serhieieva, and Z. Toporetska, "The proof of the event of a financial resources fraud in the banking sector: problematic issues," (in English), Financial and Credit Activity-Problems of Theory and Practice, Article vol. 1, no. 28, pp. 36-45, 2019.
7. K. Noor et al., "Performance analysis of a surveillance system to detect and track vehicles using Haar cascaded classifiers and optical flow method," 2017 12th IEEE Conference on Industrial Electronics and Applications (ICIEA), Siem Reap, 2017, pp. 258-263.