International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 05 Issue: 06 | June - 2021 ISSN: 2582-3930

BANKING FRAUDULENT ANALYTICS USING MACHINE LEARNING AND DEEP LEARNING APPROACH

Ritik Solanki[1], Purav Parekh[2], Raghvendra Singh Chauhan[3], Prasoon Jain[4], Keshav Purswani[3] ,

[1] Bachelor of Technology, Department of Computer Science & Engineering, Acropolis Institute of Technology and Research, Indore, Madhya Pradesh, India.
[2] Bachelor of Technology, Department of Computer Science & Engineering, Acropolis Institute of Technology and Research, Indore, Madhya Pradesh, India.
[3]Bachelor of Technology, Department of Computer Science & Engineering, Acropolis Institute of Technology and Research, Indore, Madhya Pradesh, India
[4]Bachelor of Technology, Department of Computer Science & Engineering, Acropolis Institute of Technology and Research, Indore, Madhya Pradesh, India.
[3]Assistant Professor, Department of Computer Science & Engineering, Acropolis Institute of Technology and Research, Indore, Madhya Pradesh, India.

ABSTRACT - The banking sector is a very important sector in our present-day generation where almost every human has to deal with the banks. In dealing with the banks, the customers and the banks face the chances of being trapped by fraudsters. Examples of fraud include credit card fraud, accounting fraud, etc. This paper hereby addresses bank fraud detection via the use of Machine learning, Deep learning and of data-mining techniques are clustering, forecasting, and classification to analyze the customer data in order 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. The rate at which banks loses funds to loan beneficiaries due to loan default is alarming. This trend has led to the closure of many banks, potential beneficiaries deprived of access to on; and many workers losing their jobs in the banks and other sectors. We have used past loan records to predict fraud in bank loan and subsequently avoid loan default that manual scrutiny by a credit officer would not have discovered.

KEYWORDS: Data Mining, Authentication, Real Time Fraud, Banking Sector, Deep learning, Data Sampling, Machine learning.

1.INTRODUCTION

Fraud detection is the recognition of symptoms of fraud where no prior suspicion or tendency to fraud exists. Examples include insurance fraud and accounting fraud. Data from the Nigeria Inter-Bank Settlement System (NIBSS) has revealed fraudulent transactions in the banking sector at its peak [1]. Fraud has evolved from being committed by casual fraudsters to being committed by organized crime. Any unlawful act by human beings or invoked by machines that leads to personal gain at the expense of institutions or the legal human beneficiaries is a financial fraud, but an error must not be taken for a fraud [1], [12-14]. Examples of financial fraud are money laundering, bank credit fraud, pension fraud, co-operative society fraud, telecommunications fraud, credit card fraud, inflated contracts, financial reports fraud, health insurance fraud [15], automobile insurance fraud, and mortgage insurance fraud.

2. LITERATURE REVIEW

Banks are experiencing challenges in protecting the online/internet banking channel. The challenge is in keeping the customer's account secure while avoiding complexity in the login process [5]. Fractal makes use of sophisticated analytical models and rules to correctly identify suspicious behaviour. But customers can be greatly frustrated if a transaction is incorrectly declined-meaning that banks need to select fraud prevention systems that deliver the lowest levels of false positive cases. Online Banking Authentication is getting more complex as new threats are discovered and the technology needs to secure users against them. Authentication used to be a simple password but because of growing threats over the years it has grown from that to password with numbers, then to password with numbers, symbols and special character, to knowledge-based questions and finally to the current

state of external devices (token) and communication channels to verify current occurring transactions. The solution proposed therefore is to merge the security strength of an authentication server with the logic and accuracy of a fraud detection system to identify suspect behaviour and step-up authentication, or on the other hand where normal behaviour is recognized, to step down authentication so that the customer is not unnecessarily inconvenienced – thus achieving low risk and higher customers satisfaction.

Fractals can help institutions decide:

QI. What should trigger increased or decreased authentication?

QII. Should access be allowed when high risk authentication is not available, or should restrictions be implemented?

QIII. What constitutes a "normal" profile for a customer?

In their conclusion, it was said that using fractals to enable Intelligent Risk Authentication means these strategies can be crafted, implemented and managed quickly, easily and with the end goal of greater customer satisfaction without greater risk. A. Overcoming Losses to Fraud A fraud survey conducted in 2011 by FICO states that the following is required for fraud losses to be reduced drastically [6]

1. Fraud detection in real time.
2. Analytics
3. Workflow
4. Efficient rules engine

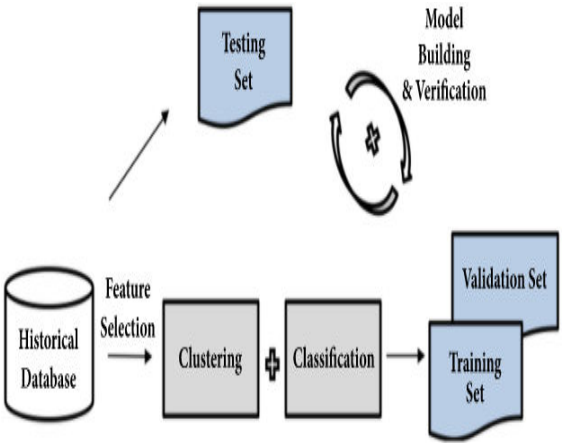
3. METHODOLOGY:

Our Banking Fraudulent Analytical System helps the banking sectors in analyzing the customers data for loan approval. In the current situation, every country's economic balance relies on the banking sector. Banking sectors depend on loan Interest to make their profits. But sometimes, due to loop-holes or slow processing in the banking system, frauds happen by people which results in loss for banks. So, Our Software will automate their slow analytic process and reduce the chances of risk in loan approval. Our Software will do in-depth analysis for customers records and predict whether the person is liable to pay the loan amount or not. The overall processes of detecting financial fraud include sampling process for class imbalance problem and feature selection process for an accurate model. The previous research papers are mainly related to specific approaches such as algorithms, which needs a further step for implementation. Moreover, the methods of machine learning, previous research only used one of the learning methods between supervised and unsupervised learning. However, our research has performed the overall process of financial fraud detection from a practical perspective based on supervised and unsupervised learning methods. Also, we are proposing a practical method by applying a sampling process and feature selection process for solving data unbalanced problems and rapid detection in the real world.

3.1. Machine Learning

Machine learning is a field where machines learn concepts using data, using statistical analysis to predict and classify and input data as an output value. The field of machine learning is divided into supervised learning and unsupervised learning depending on the learning method. Supervised learning predicts the outcome based on the values of input data provided with the given label. On the other hand, unsupervised learning is performed in a state where data is not labelled and is often called a clustering process.

The proposed model consists of data pre-processing, sampling feature selection, application of classification, and clustering algorithm based on machine learning. In this paper, the validation step is performed for each step to verify the effectiveness of the proposed financial fraud detection model. In the pre-processing process, data correlation analysis and data cleaning process which cleans the noise data are performed. The following process is the sampling process which evaluates a dataset with various ratios for verification through random oversampling and under sampling methods. After the feature selection process, the clustering process with the proposed algorithm performs and this result is used as a training set in the classification process. By applying supervised algorithms to the previous result, which was derived in the clustering process previously, the higher prediction could be achieved. The model validation process is performed with precision and recall rate through F-measure. The structure of the proposed fraud detection system model is as Figure 1.



3.2. **Sampling**

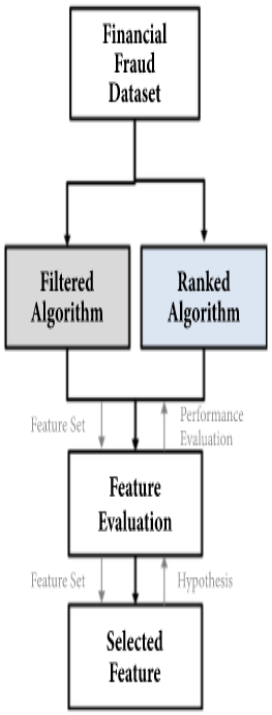
Imbalanced problems in the data could mislead the detection process to the misclassifying problem and a real transaction dataset of financial transactions usually contains a data imbalanced problem. Previous research has proposed a random minority oversampling method and clustering-based under sampling approach to select the representative data as training data to improve the classification accuracy for minority class [40].

To solve the class imbalanced problem, our research generates various datasets using Synthetic Minority Oversampling Technique (SMOTE) and Random Under sampling (RUS) for the more accurate experiment. SMOTE is an oversampling technique generating arbitrary examples rather than simply oversampling through duplication or replacement [41]. RUS was applied for downsizing the normal transactions by extracting sample data randomly for the class imbalance problem.

3.3. **Feature Selection**

Feature selection has been proven to be effective and efficient for machine learning problems. The objectives of feature selection are building simpler and more comprehensible models, improving data mining performance such as predictive accuracy and comprehensibility. The wrapper method mostly relies on the predictive performance of a predefined learning algorithm to evaluate the selected features. It repeats the searching step and evaluates criteria until desired learning performance is obtained. The drawback of the wrapper method is that the search space could be vast and it is relatively more expensive than other methods. Features are scored based on the scores according to the evaluation criteria, and the lowest scored features are removed [43]. For this reason, we applied filter-based feature selection algorithms for the feature selection method, which is the fastest and also suitable for practical use.

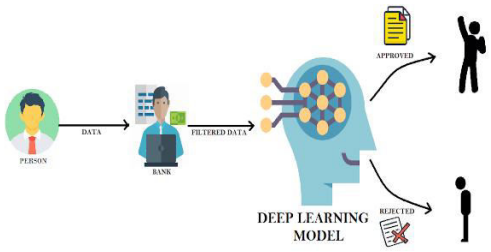
In the proposed research, we have done selection of eight subset feature selection algorithms and six ranked feature selection algorithms to select features among existing features. The results of two feature selection algorithms are combined to prioritize the features by selecting features which exceed the parameter in frequency and ranking. Figure 2 is the flowchart of the feature selection process proposed in our research.



3.4. **Deep Artificial Neural Networks**

Deep learning (DL) is a subfield of machine learning in which it works like the structure and function of the brain and called as artificial neural networks. An artificial intelligence function imitates the working of the human brain in processing data and creating patterns for use in decision-making, through the capability of unsupervised learning from data that is unstructured or unlabeled. Artificial Neural Networks (ANN) are called neural networks or multilayer perceptron's. In neural networks, the predictive capability comes from the hierarchical or multi-layered structure of the networks [45]. Also, a multilayer perceptron has a neural network with one or more intermediate layers between the input and output layers. Figure 3 is a simple artificial neural network and the middle layer between the input layer and the output layer is called a hidden layer. This network gets connected in the direction of the input layer, the hidden layer, and the output layer and is a feedforward network in which there is no direct

connection from the output layer to the input layer in each layer. Most multilayer perceptron's can be learned using backpropagation learning algorithms.

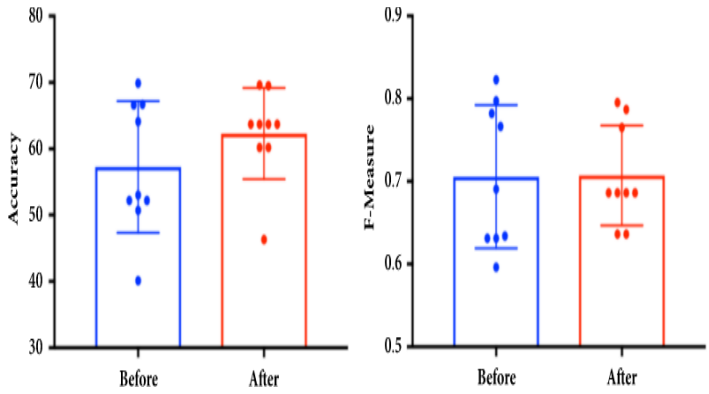


In machine learning, which is based on statistics, F-measure is a well-known measurement of model performance between predicted class and actual class using recall and precision. In our research, the F-measure is used to measure the ratio between the actual value and the value that the algorithm detects and predicts [46] and the confusion matrix used to measure the F-measure value.

4. Modelling and Analysis

I. In this paper, we aimed to discover hidden patterns by using unsupervised learning and supervised learning for more accurate classification. To design the detection system as to be useful in the operation in the real environment scenarios, we have proposed the feature selection method that can be applied to the automation system. We have constructed the system model process by applying the feature selection method on unsupervised learning algorithm firstly and then the applied supervised learning algorithm later for accurate classification based on the above experimental results by open dataset and real dataset. The final model validation was performed based on actual financial transaction data in Korea. Also, we compared the final accuracy of the proposed machine learning based detection model and the detection accuracy of models using artificial deep neural networks.

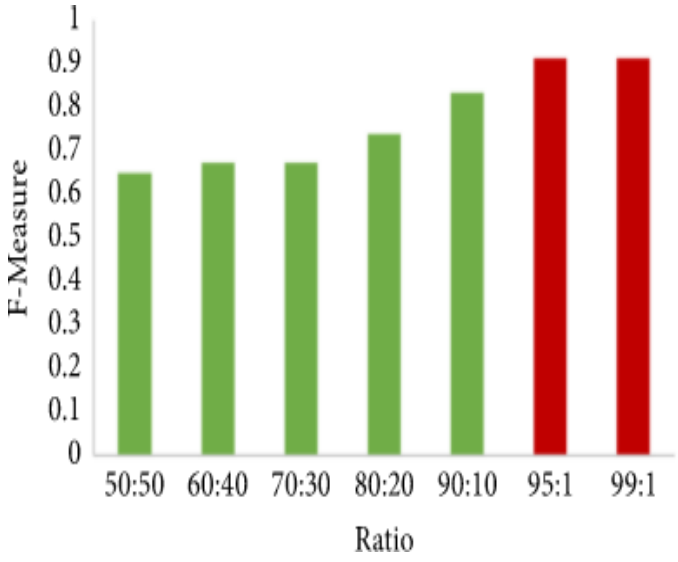
II. The machine learning based model includes various proportions of the sampling process for application in the real environment and includes an algorithm based automatic feature selection process. In addition, we apply algorithms based on unsupervised learning using selected features and apply algorithms based on supervised learning for more accurate classification.



5. Result and Discussion:

We performed the validation based on the identical actual financial transaction data for machine learning method and artificial neural network. In conclusion, the well-known machine learning method has a higher fraud detection rate than the artificial neural network. By integrating the ratios, the maximum detection rate of the machine learning method was 1, the lowest detection rate was 0.736, and the average detection rate was 0.98618 when all of the algorithms were utilized. The maximum detection rate in all ratios of the artificial neural network was 0.914, the lowest detection rate was 0.651, and the average detection rate was 0.78.

Results in Figure 1.2 show the F-measure value of the artificial neural network for detecting financial fraud in various ratios.



3. CONCLUSIONS

Several organizations experienced bank fraud to some extent. The essential thing to note is that managing fraud can be productive, and groundbreaking, and can position an association in an influential position inside its industry or business section. Adaptive data mining and intelligent analysis can play an important role in the loan fraud detection domain. They are robust enough to defeat sophisticated fraudsters, they are fast enough to minimize fraud damages, and they are scalable enough to tackle huge volumes of data. Intelligent agents will eventually be the ultimate means to fight against loan frauds. However, there is still a long way to go before the wide adoption of intelligent agents for loan fraud detection. The accuracy of fraud detection needs to be improved; the reliability of the agents needs to be ensured by testing the system on a real bank server to check its performance and acceptability.

ACKNOWLEDGEMENT

We sincerely thanks to our Professor Keshav Purswani and professor Kavita Namdev for their support and guidance throughout the whole work.

REFERENCES:

[1]. Baruah, S.K. (2015). RBI chief Wants PMO to Act against Bank Frauds Worth Rs. 17,500 crores, The Hindustan Times, April 24, available at www.hindustantimes.com.

[2]. Bhasin, M. L. (2012). Audit Committee Scenario and Trends in a Developing Country, School of Doctoral Studies European Union Journal, 4, 53-70.

[3]. Bhasin, M. L. (2013). Corporate Governance and Forensic Accountant: An Exploratory Study, Journal of Accounting, Business and Management, October, 20(2), 55-75.

[4]. Bhasin, M. L. (2015). Menace of Frauds in the Indian Banking Industry: An Empirical Study, Australian Journal of Business and Management Research, 4(2), April, 21-33.

[5]. Bhasin, M. L. (2016). Contribution of Forensic Accounting to Corporate Governance: An Exploratory Study of an Asian Country, International Business Management, 10(4), 479-492. (Forthcoming)