International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 10 | October - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

INTELLIGENT FRAUD DETECTION IN FINANCIAL SYSTEM USING MACHINE LEARNING TECHNIQUE

Mrs.N.SUGANTHI [1] , Mrs.S.P.AUDLINE BEENA[2]

PG Scholar[1], Department of Computer Science, SRI Muthukumaran Institute of Technology, Chennai
Assistant Professor[2], Department of Computer Science, SRI Muthukumaran Institute of Technology, Chennai

ABSTRACT

Fraudulent financial statements (FFS) are the results of manipulating financial elements by overvaluing incomes, assets,sales, and profits while underrating expenses, debts, or losses. To identify such fraudulent statements, traditional methods, including manual auditing and inspections, are costly, imprecise, and time-consuming. Intelligent methods can significantly help auditors in analyzing a large number of financial statements. In this study, we systematically review and synthesize the existing literature on intelligent fraud detection in corporate financial statements. In particular, the focus of this review is on exploring machine learning and data mining methods, as well as the various datasets that are studied for detecting financial fraud. We adopted the Kitchenham methodology as a well-defined protocol to extract, synthesize, and report the results. Accordingly, 47 articles were selected, synthesized, and analyzed. We present the key issues, gaps, and limitations in the areaof fraud detection in financial statements and suggest areas for future research. Since supervised algorithms were employed more than unsupervised approaches like clustering, the future research should focus on unsupervised, semi-supervised, as well as bio-inspired and evolutionary heuristic methods for anomaly (fraud) detection. In terms of datasets, it is envisaged that future research making use of textual and audio data. While imposing new challenges, this unstructured data deserves further study as it can show interesting results for intelligent fraud detection.

INTRODUCTION

Fraud can be defined as a case where a person uses someone else's credit card for personal reasons while the owner and the card issuing authorities are unaware of the fact that the card is being used. Fraud detection involves monitoring the activities of populations of users in order to estimate, perceive or avoid objectionable behavior, which consist of fraud, intrusion, and defaulting.

The Bidder shall extend their system or provide a dedicated dashboard/portal to Bank for generating real time MIS reports as required by the Business. The portal shall also be integrated with Bidder's host / Authorization system for providing the services likeTransaction details, Transaction status, Card status, List of Cards issued, Billing Statements related etc.

Bidder's Card Management System shall be interfaced with Bank's Card Origination System for the generation of Cards throughStraight Through Processing (STP). For example, the bank may intend to issue instant virtual Credit Cards to Preapproved Category of Customers. Bank's Card Origination System will expose its interface to customers for getting their inputs. On successful validation of Card Origination System, the information will flow to Bidder's Card Management System which shallgenerate the Cards and to send the response back to the customer. In an ideal scenario the entire workflow shall be completed within 2 minutes.

EXISTING SYSTEM

This project of existing system, a research about a case study involving financial statement fraud detection, where data normalization is applied before analysis and with results obtained from the use of artificial neural networks on fraud detection. The promising results can be obtained by using normalized data and data should be trained. This research was based on unsupervised learning. Significance of this paper was to find new methods for fraud detection and to increase the accuracy of results.

DISADVANTAGES

- In this paper a new collative comparison measure that reasonably represents the gains and losses due to fraud detection isproposed.
- A cost sensitive method which is based on Bayes minimum risk is presented using the proposed cost measure.

PROPOSED SYSTEM:

This project is proposed system, we are applying random forest algorithm for classification of the financial statement dataset.

Random forest is an algorithm for classification and regression. The proposed model outperforms the state-of-the-art machine learning and data mining algorithms for financial statement detection problems. In addition, we have performed experiments by balancing the data and applying algorithms to minimize the false negative rate. The proposed approaches can be implemented effectively for the real-world detection of financial statement fraud. The machine learning model is proposed by adding a number ofadditional layers that are then used to extract the features and classification from the financial statement fraud detection dataset. To perform a comparative analysis between ML with DM algorithms and proposed with baseline model, the results prove that the proposed approach outperforms existing approaches.

SYSTEM ARCHITECTURE:

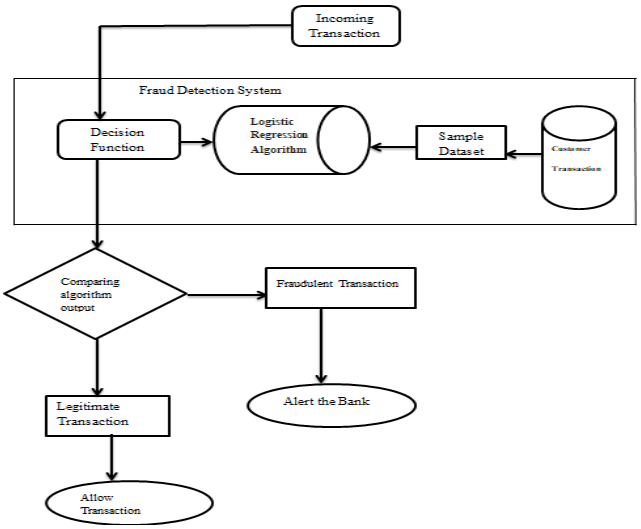


Fig . Architecture of this system

ADVANTAGES:

- It is used to detect the fraud transaction.
- Data imbalance problem is rectified.

ALGORITHMS USED WITH PSEUDOCODE:

Regression S Given the bioactivity vectors for all targets, x1. . . xm ∈ Rn, and the size of informer set nA; Split the data into 5 folds, each fold with roughly the same number of targets;

for K = 1, . . . , 5 do

Take the K-th fold of the data as the test data, and the rest as the training set;

for j = 1, . . . , n; .

do

Linearly scale the features such that (xi)j for all i in the training set lie in the range [0, 1];

end

for k = 2, 3, . . . do

Cluster the training data to k categories using kmeans++ with 100 repeats;

Select the informer set A with nA features by the greedy heuristic based on the regularized logistic regression model ; Train a

new logistic regression model (5) using the selected coordinates A; Use the logistic regression model to predict on

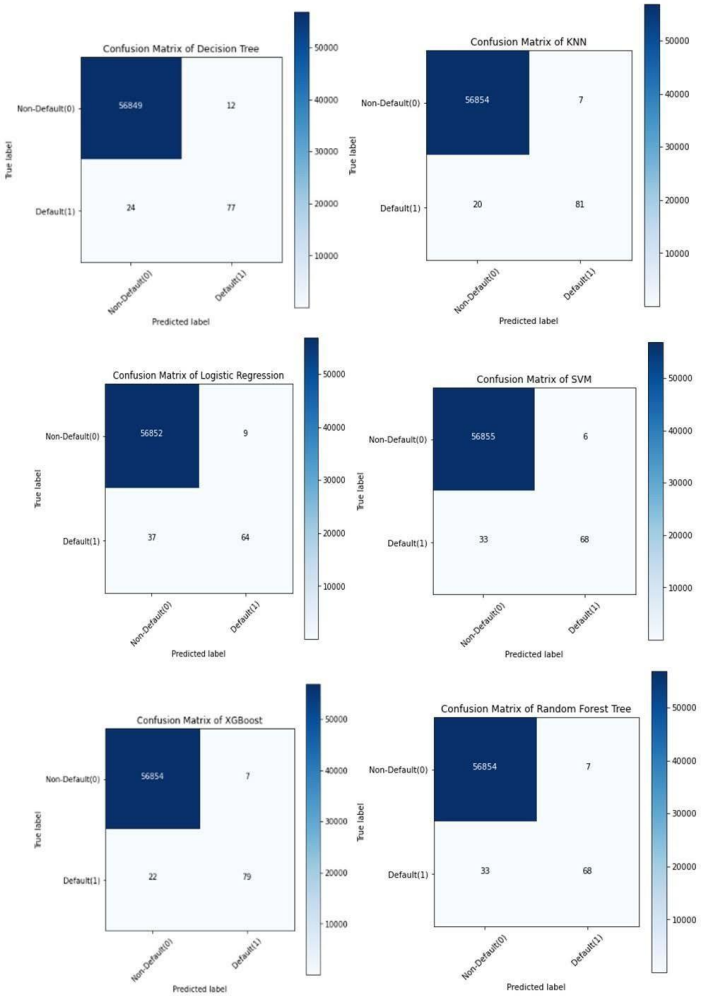
the test set through and evaluate the performance;

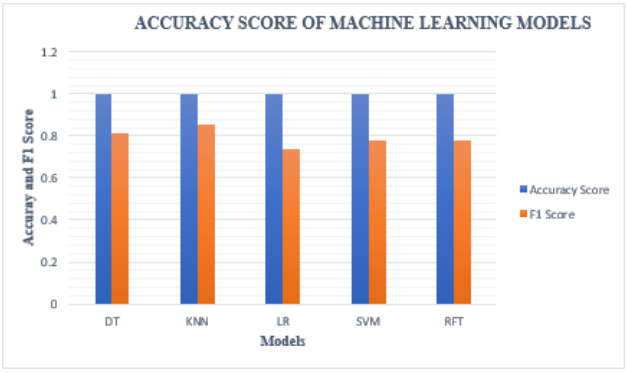
end

end

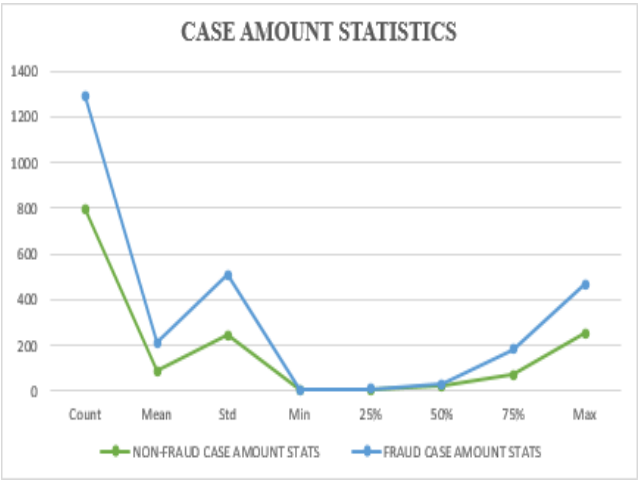
Rescale the whole data set just as in the cross validation procedure; Use the best k selected by cross validation to cluster the data; Use the greedy heuristic to select the informer set A with size n_A; Train the logistic regression model on the whole informer set Awith all target

Result Analysis





As shown in Figure the case count statistics, the valuesof the 'Amount' variable vary substantially once associated with the respite of the variables. To decrease the wide rangeof the values, we can standardize it by means of the *'Standard-Scaler'* method in Python.



CONCLUSION

Financial statement fraud detection FSFD is a developing area in which it is advantageous to outrun the fraudsters. Besides, there are still aspects of intelligent FSFD that have not been investigated thoroughly. In this project performed a systematic literature review using Kitchenham methodology to analyze the FSFD problem in terms of machine learning/data mining approaches and datasets used in the studies. It presents some of the key issues, gaps, and limitations of FSFD and suggested future research areas. Our study presented various ML/DM algorithms employed in the existing literature. Categorizing FSFD methods by the ML/DM fraud detection technique is an efficient method to identify the promising practices for this area of research. In addition to exploring the ML/DM techniques, this study focuses on analyzing the datasets used for financial fraud detection in the existing literature.

REFERENCES:

1. "M. C. M. Oo and T. Thein, "An efficient predictive analytics system for high dimensional big data", *J. KingSaud Univ. Comput. Inf. Sci.*, vol. 34, no. 1, pp. 1521-1532, Jan. 2022.

2. S. A. Ebiaredoh-Mienye, E. Esenogho and T. G. Swart, "Artificial neural network technique for improving prediction of credit card default: A stacked sparse autoencoder approach", *Int. J. Electr. Comput. Eng.*, vol. 11, no. 5, pp. 4392, Oct. 2021.

3. V. Arora, R. S. Leekha, K. Lee and A. Kataria, "Facilitating user authorization from imbalanced data logs ofcredit cards using artificial intelligence", *Mobile Inf. Syst.*, vol. 2020, pp. 1-13, Oct. 2020.

4. J. Błaszczyński, A. T. de Almeida Filho, A. Matuszyk, M. Szeląski and R. Słowiński, "Auto loan fraud detection using dominance-based rough set approach versus machine learning methods", *Expert Syst. Appl.*, vol. 163, Jan. 2021.

5. S. S. Lad, I. and A. C. Adamuthe, "Malware classification with improved convolutional neural network model", *Int. J. Comput. Netw. Inf. Secur.*, vol. 12, no. 6, pp. 30-43, Dec. 2021.

6. Benchaji, S. Douzi and B. E. Ouahidi, "Credit card fraud detection model based on LSTM recurrent neural networks", *J. Adv. Inf. Technol.*, vol. 12, no. 2, pp. 113-118, 2021.

7. Thennakoon, C. Bhagyani, S. Premadasa, S. Mihiranga and N. Kuruwitaarachchi, "Real-time credit card fraud detection using machine learning", *Proc. 9th Int. Conf. Cloud Comput. Data Sci. Eng. (Confluence)*, pp. 488-493, Jan. 2019.

8. D. Elreedy and A. F. Atiya, "A comprehensive analysis of synthetic minority oversampling technique (SMOTE) for handling class imbalance", *Inf. Sci.*, vol. 505, pp. 32-64, Dec. 2019.

9. J. Gao, Z. Zhou, J. Ai, B. Xia and S. Coggeshall, "Predicting credit card transaction fraud using machine learning algorithms", *J. Intell. Learn. Syst. Appl.*, vol. 11, pp. 33-63, 2019.

10. Eshghi and M. Kargari, "Introducing a method for combining supervised and semi-supervised methods in fraud detection", *Proc. IIIEC*, pp. 23-30, Jan.