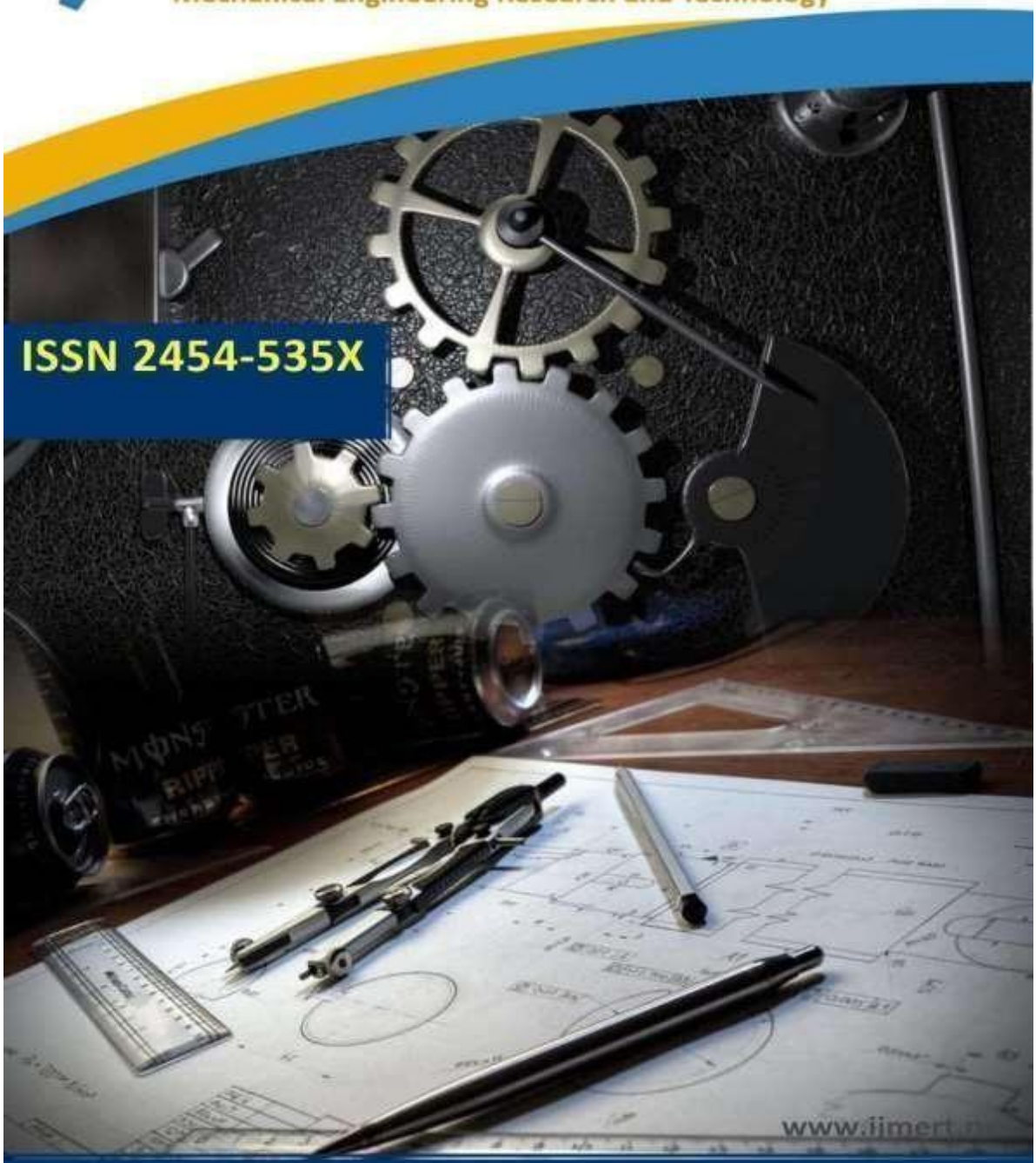
International Journal of Mechanical Engineering Research and Technology



ISSN 2454-535X

Email ID: info.ijmert@gmail.com or editor@ijmert.net

International Journal of Mechanical Engineering Research and Technology
ISSN 2454 – 5X

http://www.ijmert.com
Vol.16 Issue2, 2024

FRAUD DETECTION IN BANKING DATA BY MACHINE LEARNING TECHNIQUES

[1]MR. B. SRAVAN KUMAR,[2]KUPPILI SAI SURYA,[3]NAGULA BHAVANA,[4]ANNADI KALYAN,[5]THUMAR JAYA NARSHIBHAI,[6]MOHAMAD ABDUL RAHEEM ARIF

[1]Assistant Professor,department of information technology, malla reddy institute of engineering and technology(autonomous),dhulapally,secundrabad,bandashravan09@gmail.com

[2,3,4,5,6]UG students, department of information technology,malla reddy institute of engineering and technology(autonomous),Dhulapally,Secundrabad

ABSTRACT

As technology advanced and e-commerce services expanded, credit cards became one of the most popular payment methods, resulting in an increase in the volume of banking transactions. Furthermore, the significant increase in fraud requires high banking transaction costs. As a result, detecting fraudulent activities has become a fascinating topic. In this study, we consider the use of class weight-tuning hyperparameters to control the weight of fraudulent and legitimate transactions. We use Bayesian optimization in particular to optimize the hyperparameters while preserving practical issues such as unbalanced data. We propose weight-tuning as a pre-process for unbalanced data, as well as CatBoost and XGBoost to improve the performance of the LightGBM method by accounting for the voting mechanism. Finally, in order to improve performance even further, we use deep learning to fine-tune the hyperparameters, particularly our proposed weight-tuning one. We perform some experiments on real-world data to test the proposed methods. To better cover unbalanced datasets, we use recall-precision metrics in addition to the standard ROC-AUC. CatBoost, LightGBM, and XGBoost are evaluated separately using a 5-fold cross-validation method. Furthermore, the majority voting ensemble learning method is used to assess the performance of the combined algorithms. LightGBM and XGBoost achieve the best level criteria of ROC-AUC D 0.95, precision 0.79, recall 0.80, F1 score 0.79, and MCC 0.79, according to the results. By using deep learning and the Bayesian optimization method to tune the hyperparameters, we also meet the ROC-AUC D 0.94, precision D 0.80, recall D 0.82, F1 score D 0.81, and MCC D 0.81. This is a significant improvement over the cutting-edge methods we compared it to.

International Journal of Mechanical Engineering Research and Technology
ISSN 2454 – 5X
http://www.ijmert.com
Vol.16 Issue2, 2024

I.INTRODUCTION

The "Fraud Detection in Banking Data by Machine Learning Techniques" project is a critical initiative aimed at addressing the growing threat of fraudulent activities in the banking sector. With the rapid digitization of financial services and the increasing volume of transactions conducted online, detecting fraudulent behavior has become a paramount concern for financial institutions. Traditional methods of fraud detection often rely on rule-based systems that struggle to keep pace with the evolving tactics of fraudsters. In response to this challenge, this project proposes the utilization of machine learning techniques to analyze banking data and identify patterns indicative of fraudulent transactions. By leveraging advanced algorithms and models, such as logistic regression, random forests, and neural networks, the project aims to develop a robust fraud detection system capable of accurately distinguishing between legitimate and fraudulent activities. Through the implementation of machine learning-based approaches, financial institutions can enhance their ability to detect and prevent fraud, safeguarding the interests of both customers and stakeholders.

II.EXISTING SYSTEM

Halvaiee&Akbari study a newmodel called the AIS-based fraud detection model (AFDM). They use the Immune System Inspired Algorithm (AIRS) to improve fraud detection accuracy. The presented results of their paper show that their proposed AFDM improves accuracy by up to 25%, reduces costs by up to 85%, and reduces system response time by up to 40% compared to basic algorithms [11].

Bahnsen et al. developed a transaction aggregation strategy and created a new set of features based on the periodic behaviour analysis of the transaction time by using the von Mises distribution. In addition, they propose a newcost-based criterion for evaluating credit card fraud detection's models and then, using a real credit card dataset, examine how different feature sets affect results. More precisely, they extend the transaction aggregation strategy to create newoffers based on an analysis of the periodic behaviour of transactions [12]. Randhawa et al. study the application of machine learning algorithms to detect fraud in credit cards. They _rst use Naïve Bayes, stochastic forest and decision trees, neural networks, linear

International Journal of Mechanical Engineering Research and Technology
ISSN 2454 – 5X
http://www.ijmert.com
Vol.16 Issue2, 2024

regression (LR), and logistic regression, as well as support vector machine standard models, to evaluate the available datasets. Further, they propose a hybrid method by applying AdaBoost and majority voting. In addition, they add noise to the data samples for robustness evaluation. They perform experiments on publicly available datasets and show that majority voting is effective in detecting credit card fraud cases [6].

Porwal and Mukund propose an approach that uses clustering methods to detect outliers in a large dataset and is resistant to changing patterns [13]. The idea behind their proposed approach is based on the assumption that the good behavior of users does not change over time and that the data points that represent good behaviour have a consistent spatial signature under different groupings. They show that fraudulent behaviours can be detected by identifying the changes in this data. They show that the area under the precision-recall curve is better than ROC as an evaluation criterion [13]. The authors in [14], propose a group learning framework based on partitioning and clustering of the training set. Their proposed framework has two goals: 1) to ensure the integrity of the sample features, and 2) to solve the high imbalance of the dataset. The main feature of their proposed framework is that every base estimator can be trained in parallel, which improves the effectiveness of their framework. Itoo et al. use three different ratios of datasets and an oversampling method to deal with the problem of data imbalance. Authors use three machine learning algorithms: logistic regression, Naive Bayes, and K-nearest neighbor. The performance of the algorithms is measured based on accuracy, sensitivity, specificity, precision, F1-score, and area under the curve. They show that the logistic regression-based model outperforms the other commonly used fraud detection algorithms in the paper [15].

The authors in [16] propose a framework that combines the potential of meta-learning ensemble techniques and a cost sensitive learning paradigm for fraud detection. They perform some evaluations, and the results obtained from classifying unseen data show that the cost-sensitive ensemble classifier has acceptable AUC value and is efficient as

International Journal of Mechanical Engineering Research and Technology
ISSN 2454 – 5X
http://www.ijmert.com
Vol.16 Issue2, 2024

compared to the performances of ordinary ensemble classifiers. Altyeb et al. propose an intelligent approach for detecting fraud in credit card transactions [17]. Their proposed Bayesian-based hyperparameter optimization algorithm is used to tune the parameters of a LightGBM. They perform experiments on publicly available credit card transaction datasets. These datasets consist of fraudulent and legitimate transactions. Their evaluation results are reported in terms of accuracy, area under the receiver operating characteristic curve (ROC-AUC), precision, and F1-score metrics.

Xiong et al. propose a learning-based approach to tackle the fraud detection problem. They use feature engineering techniques to boost the proposed model's performance. The model is trained and evaluated on the IEEE-CIS fraud dataset. Their experiments show that the model outperforms traditional machine-learning-based methods like Bayes and SVM on the used dataset [18].Viram et al. evaluate the performance of Naive Bayes and voting classifier algorithms. They demonstrate that in terms of evaluated metrics, particularly accuracy, the voting classifier outperforms the Naive Bayes algorithm [19].

Disadvantages

- The system never use a sequential model, which is a linear stack of layers to construct an artificial neural network model. Our model has a dense class, which is a very common layer and is often used.
- The system never implements Majority Voting model which leads less effective.

III.PROPOSED SYSTEM

The system proposes an efficient approach for detecting credit card fraud that has been evaluated on publicly available datasets and has used optimized algorithms SVM and logistic regression individually, as well as majority voting combined methods, as well as deep learning and hyper parameter settings. An ideal fraud detection system should detect more fraudulent cases, and the precision of detecting fraudulent cases should be high, i.e., all results should be correctly detected, which will lead to the trust of customers in the bank, and on the other hand, the bank will not suffer losses due to incorrect detection. propose a group

International Journal of Mechanical Engineering Research and Technology
ISSN 2454 – 5X
http://www.ijmert.com
Vol.16 Issue2, 2024

learning framework based on partitioning and clustering of the training set. Their proposed framework has two goals: 1) to ensure the integrity of the sample features, and 2) to solve the high imbalance of the dataset. The main feature of their proposed framework is that every base estimator can be trained in parallel, which improves the effectiveness of their framework.

Advantages

_ We adopt Bayesian optimization for fraud detection and propose to use the weight-tuning hyperparameter to solve the unbalanced data issue as a pre-process step. We also suggest using CatBoost and XGBoost alongside LightGBM to improve performance. We use the XGBoost algorithm due to the high speed of training in big data as well as the regularization term, which overcomes overfitting by measuring the complexity of the tree, and it does not require much time to set the hyper parameters. We also use the Catboost algorithm because there is no need to adjust hyper parameters for overfitting control, and it also obtains good results without changing hyper parameters compared to other machine learning algorithms.

_ We propose a majority-voting ensemble learning approach to combine CatBoost, XGBoost, and Light- GBM and review the effect of the combined methods on the performance of fraud detection on real, unbalanced data.We also propose to use deep learning for adjusting and _ne-tuning the hyper parameters.

_ To evaluate the performance of the proposed methods, we perform extensive experiments on real-world data. To better cover the unbalanced datasets, we use recall precision in addition to the typically used ROC-AUC. We also evaluate the performance using F1_score and MCC metrics. According to the results, the proposed methods outperform the existing and based methods. For evaluations, we use publicly available datasets and also publish the source codes 1 with public access to be used by other researchers.

IV.MODULES

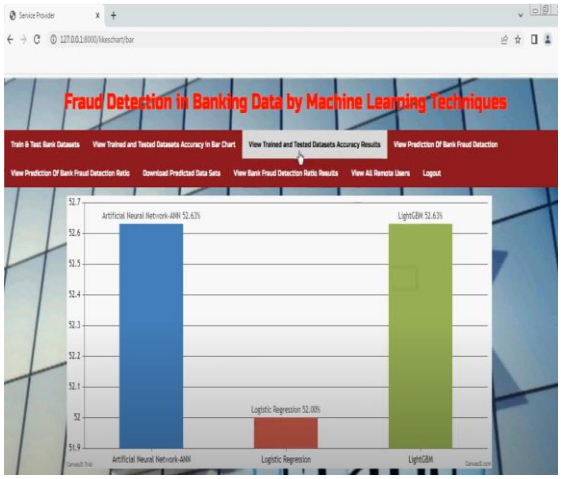
- **Service Provider**

In this module, the Service Provider has to login by using valid user name and password.

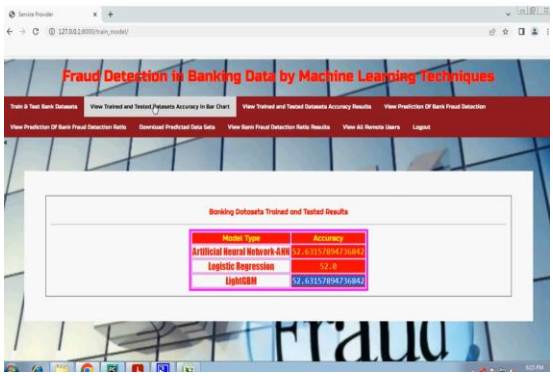
International Journal of Mechanical Engineering Research and Technology
ISSN 2454 – 5X

http://www.ijmert.com
Vol.16 Issue2, 2024

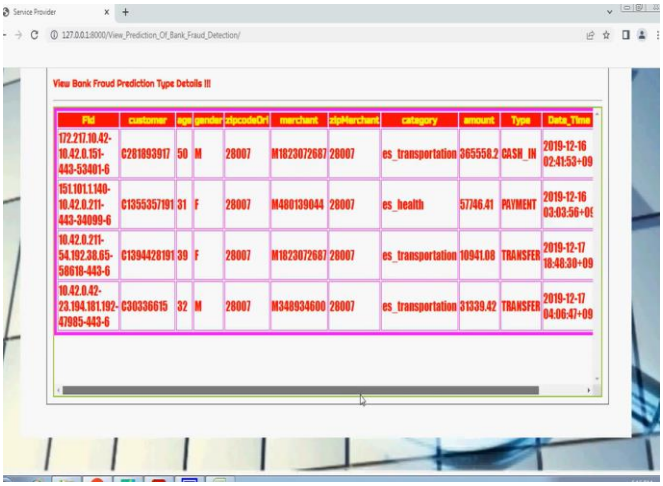
After login successful he can do some operations such as Browse Datasets and Train & Test Data Sets, View Trained and Tested Accuracy in Bar Chart,



View Trained and Tested Accuracy Results,



View Predicted Type, View Type Ratio, Download Predicted Data Sets,



View Type Ratio Results, View All Remote Users.

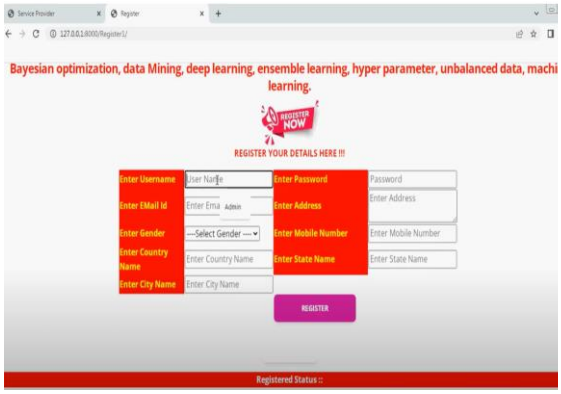
- **View and Authorize Users**

In this module, the admin can view the list of users who all registered. In this, the admin can view the user's details such as, user name, email, address and admin authorizes the users.

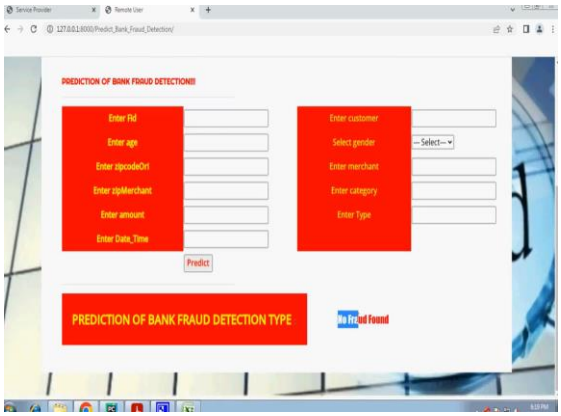
- **Remote User**

In this module, . User should register before doing any operations. Once user registers, their details will be stored to the database.

International Journal of Mechanical Engineering Research and Technology
ISSN 2454 – 5X
http://www.ijmert.com
Vol.16 Issue2, 2024



After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like REGISTER AND LOGIN, after login we have to Predict Type,



VIEW YOUR PROFILE.

V.CONCLUSION

In conclusion, the "Fraud Detection in Banking Data by Machine Learning Techniques" project holds immense promise for enhancing fraud detection capabilities in the banking sector. By harnessing the power of machine learning techniques, financial institutions can improve the accuracy and efficiency of their fraud detection systems, thereby reducing financial losses and protecting against reputational damage. The development of robust machine learning models capable of analyzing complex banking data and identifying suspicious patterns is essential for staying ahead of increasingly sophisticated fraud schemes. Through the deployment of machine learning-based fraud detection systems, financial institutions can strengthen their security posture, instill trust among customers, and uphold the integrity of the banking system as a whole.

VI.REFERENCES

- Bhattacharyya, S., & Jha, S. (2018). Fraud Detection in Banking Using Machine Learning Techniques: A Review. In 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON) (pp. 0132-0137). IEEE.
- Dal Pozzolo, A., Boracchi, G., Caelen, O., Alippi, C., & Bontempi, G. (2015). Credit card fraud detection: a realistic modeling and a novel learning strategy. IEEE Transactions on Neural Networks

International Journal of Mechanical Engineering Research and Technology
ISSN 2454 – 5X
http://www.ijmert.com
Vol.16 Issue2, 2024

and Learning Systems, 29(8), 3784-3797.

- Deng, X., Li, Y., & Pan, S. (2016). Application of machine learning in bank credit card fraud detection. In 2016 35th Chinese Control Conference (CCC) (pp. 9510-9513). IEEE.
- Phua, C., Lee, V., Smith, K., & Gayler, R. (2005). A comprehensive survey of data mining-based fraud detection research. arXiv preprint cs/0412098.
- Ribeiro, F., Mariani, V., Ribeiro, M., & Ruas, T. (2017). Fraud detection in banking transactions: A machine learning approach. In 2017 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC) (pp. 104-111). IEEE.
- Santhi, G. (2017). Fraud detection using machine learning techniques. International Journal of Advanced Research in Computer Science, 8(2), 367-370.
- Srivastava, R. (2018). Machine Learning Algorithms for Credit Card Fraud Detection. In Advances in Data and Information Sciences (pp. 163-173). Springer, Singapore.
- Taylor, R. C. (2009). An overview of credit card fraud detection techniques: 2010-2014. International Journal of Computer Applications, 3(1), 38-46.
- Wang, H., & Lu, Y. (2003). A credit card fraud detection model based on neural network with feature selection. In Proceedings of the International Joint Conference on Neural Networks (IJCNN'03) (Vol. 3, pp. 1937-1941). IEEE.
- Wu, C. S., & Ting, T. O. (2014). An effective method to detect credit card fraud by using rough set theory and neural network. Expert Systems with Applications, 41(5), 2475-2482.