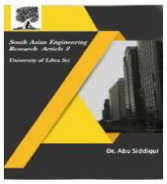
International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal



COST-SENSITIVE PAYMENT CARD FRAUD DETECTION BASED ON DYNAMIC RANDOM FOREST AND K-NEAREST NEIGHBORS

Mr.K.Shouryadhar[1], Erpula.Geetanjali[2], G.Shruthi[3], Karri Lakshmi Sai Sravya nikitha[4]

[1]Assistant Professor, School of CSE,Malla Reddy Engineering College For Women (UGC-Autonomous), Maisammaguda,Dhulapally,Secunderabad,Telangana-500100

[234]UG Student, School of CSE,Malla Reddy Engineering College for Women, (UGC-Autonomous), Maisammaguda,Dhulapally,Secunderabad,Telangana-500100

Email: shourya.ravi@gmail.com

ABSTRACT

Increasing usage of electronic payment systems has made card fraud a huge concern for cardholders and financial institutions alike. Most traditional methods of fraud detection thrive to maximize the accuracy while being insensitive to the differential cost of false positives - the good transactions wrongly flagged - and false negatives - the frauds not detected. It advocates a cost-sensitive approach to the detection of card fraud integrating Dynamic Random Forest DRF and K -Nearest neighbors algorithms in order to improve fraud detection accuracy by avoiding cost for misclassifications.One of the major challenges in effective fraud detection is the extraction of relevant features from transaction data. Cardholder spending behaviors do change over time, making more recent transactions more predictive of typical activity for a cardholder. The approach proposed here assigns greater weight to the more recent transactions, letting the model adapt to changed behavior patterns.The DRF algorithm is used in classifying transactions as genuine or forged, which grows with every transaction data, whereas the KNN algorithm supports the finding of local patterns by considering similarities that exist between two transactions. The combination of the two algorithms enables the system to pick up the global and the local spending patterns, and this reinforces the overall process of detecting.The proposed cost-sensitive approach aims at achieving cost-effective misclassifications that balance detection accuracy against the cost. This helps deliver an efficient fraud-detecting system that enhances decision making through better protection of financial consumers as well as financial institutions against fraud.

Keywords: Payment card , Fraud Detection, Dynamic Random Forest , K-Nearest Neighbor Algorithm

I. INTRODUCTION

The rapid advancement of technology in the financial sector has led to a significant rise in the use of electronic payment systems, making card transactions an essential part of global commerce. With the convenience of these systems, however, comes the increasing risk of fraudulent activities. Payment card fraud, including unauthorized transactions, identity theft, and other deceptive practices, has become a major challenge for financial institutions, merchants, and consumers alike. As sophistication in fraudsters also requires a

change in the fraud detection system to accommodate new techniques of fraudulent activity.Rule-based systems and manual intervention are mostly used in traditional fraud detection methods. The methods cannot handle increased complexity and volume of fraudulent transactions. One of the most important limitations of conventional fraud detection systems is their single-minded focus on accuracy without considering asymmetrical costs of misclassifications. The cost to financial institutions differs from that of cardholders both in false positives referring to legitimate transactions wrongly identified as fraudulent and false negatives which are fraudulent transactions not detected. For example, false positives result in customer dissatisfaction and false negatives lead to a substantial loss of money.

This proposed work will overcome all these difficulties with the cost-sensitive approach to card fraud detection, where algorithms such as Dynamic Random Forest (DRF) and k-nearest neighbors (KNN) have been applied. The DRF algorithm is a form of ensemble learning, which is perfectly suitable to hold the dynamic nature of fraud. It can create several decision trees, each trained on a different subset of transaction data. This method enhances the capability of the model to discover subtle patterns in the data that would point towards fraudulent behavior. The KNN algorithm, on the other hand, focuses on anomalies in the transaction data, which classify transactions based on the majority class of their nearest neighbors. This pattern recognition at a local level boosts subtle fraud patterns that the global models cannot detect.

This is done with a cost-sensitive framework so that it will focus on those transactions having higher costs of misclassification in order to effectively balance out the trade-offs between detection of fraud and minimal customer disruption. This paper also solves the problem of dataset imbalance, where the fraudulent transactions are much fewer than legitimate ones, which might lead to biased model training if this problem is not appropriately addressed.

The proposed system will attempt to combine DRF and KNN for improved detection accuracy, reduction of false positives and false negatives, and providing an adaptive solution that can evolve with changing fraud patterns. This project is mainly fraud detection with minimal financial impact in terms of errors, hence a cost-effective solution for financial institutions. This helps to enhance the pervasive efforts toward protecting electronic payment systems against losses and risks caused by frauds that infringe on credit card payments.

II. RELATED WORK

2014-A. C. Bahnsen et al. The authors introduced a credit card fraud detection approach, which had the calibration of probabilities and Bayes Minimum Risk. This proposed method attempted to enhance accuracy in detection through decision threshold regulation, thus eliminating misclassification error. Their technique targeted the improvement of reliability in fraud detection systems by balancing the occurrence of false positives and false negatives.

2015 - M. Zareappor & P. Shamsolmoali used a Bagging ensemble classifier that combined multiple classifiers, including Naive Bayes (NB), Support Vector Machines (SVM), and K-Nearest Neighbor (KNN). The authors showed that ensemble methods could improve fraud detection by taking advantage of the

International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

strengths of each algorithm, making the system more robust in identifying fraudulent transactions despite variations in transaction patterns.

2017 - J. O. Awoyemi et al. Awoyemi et al. presented a comparative analysis of various techniques for machine learning to be applied for credit card fraud detection. Here, a GA is applied in conjunction with DT, RF, and LR algorithms for feature selection. Their study focused on the aspect of feature selection to enhance the performance of fraud detection models. In this case, Random Forest showed the best result because it could handle large datasets and capture complex relationships in the data.

2019 - B. Meenakshi Devi et al. authors focused on the use of Random Forest for credit card fraud detection. They applied classification and regression algorithms to show that Random Forest is highly effective in identifying fraudulent transactions due to its ensemble learning mechanism, which combines the outputs of multiple decision trees. This approach improved the detection of fraudulent transactions and minimized the likelihood of misclassifications.

The introduction The system architecture begins with data collection, where the datasets are split into features (inputs) and labels (outputs). Data is further split into training and test sets.

In the feature engineering stage, raw data is processed and appropriate features are selected to optimize model performance. A variational auto-encoder (VAE) is used for dimensionality reduction, while a heterogeneous random forest (HRF) processes the selected features for robust classification and prediction.

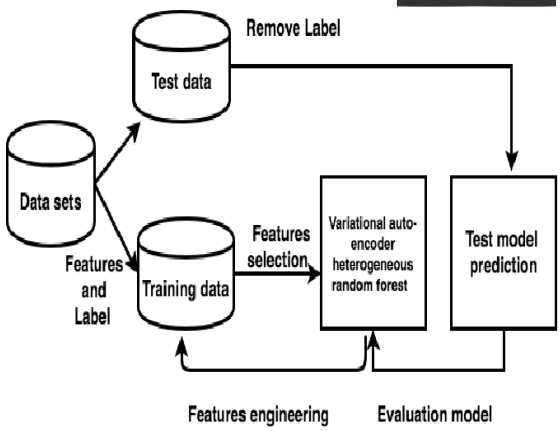


Fig:1 : System Architecture

The model is tested using test data for accuracy and reliability. The final deployment of the model leads to prediction outcomes, which presents an efficient and accurate system to detect patterns or anomalies.

III. IMPLEMENTATION

Implementation steps of the proposed cost-sensitive credit card fraud detection system are given below:

1.Environment Setup:

The setup computation environment is the first step for implementing this system. It will code the system using Python and packages for manipulating the data as well as implementation algorithms are Pandas, NumPy, Scikit-learn, and Matplotlib. More intricate model building in specific cases will be helpful through TensorFlow and PyTorch.

2. Data Preprocessing:

Loading the data set and inappropriatehandling of missing or erroneous values.Scaling of features like normalization to have the same scale for all features Class imbalance techniques in which methods like Synthetic Minority Oversampling Technique (SMOTE)

are used to resample the minority class in this case, fraudulent transactions

3. Feature Engineering:

Select features based on relevance to detect fraud cases. Use techniques in dimensionality, like PCA or VAE, that reduce to only the most relevant attributes of interest for the model

4. Modeling/Developing

DRF: A DRF model-classify transaction with the combination of a decision tree through dynamic pattern adjustments due to transactions.

K-Nearest Neighbors (KNN): This algorithm classifiestransactions by how similar they are to known patterns of fraud in the feature space.

5. Cost-Sensitive Training:

The models are trained in a cost-sensitive manner. Here, there is a larger penalty for false negatives, missed fraudulent transactions. That will prevent financial loss due to misclassification.

6. Model Integration:

An ensemble learning technique can be applied by combining the models of DRF and KNN, so that final prediction is made using a weighted majority vote or other fusion techniques.

7. Evaluation:

The system will be evaluated on the test dataset against metrics such as precision, recall, F1-score, and cost efficiency. The cost-sensitive metric ensures further economic viability of the model.

8. Deployment:

The trained model goes into a real-time pipeline of fraud detection through APIs or directly implemented to the financial systems for practicing transaction monitoring.

This systematic implementation ensures a robust solution, scalable and adaptive which detects fraudulent transactions.

IV. ALGORITHM

The Dynamic Random Forest (DRF) and K-Nearest Neighbors (KNN) algorithm that uses payments card fraud detection method involves an orderly process to detect the fraudulent transactions.

Step 1: Collect and Preprocess Data

The algorithm begins with data collection for the transactions, including features like transaction ID, amount, time, and more about the merchant. Data cleaning was performed to deal with missing values, and the dataset was split into training and testing sets, class imbalance being handled through techniques like oversampling or undersampling.

Step 2: Feature Extraction

Extracting the features of key characteristics in the process, such as frequency of transactions, average spend, time of transaction, and location. These features are normalized to the standard scale, and the feature selection is carried out by retaining the most informative attributes for model training.

Step 3: Training of the Model using DRF

The Dynamic Random Forest (DRF)learns from the dataset. DRF is an ensemble approach which makes multiple decision trees where every tree focuses on learning the patterns in the data set differently. It is helpful in large data sets and when the classes are imbalanced. DRF examines the hierarchical structure to differentiate fraudulent from legitimate

International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

transactions, which can increase the detection accuracy and decrease the false positives.

Step 4: Train a model on KNN

At the same time, it has a K-Nearest Neighbors (KNN) algorithm used to classify transactions based on their proximity to `K` nearest neighbors in the feature space. Thus, it identifies some kind of local patterns and anomalies missed by DRF.

Step 5: Prediction Combine

The transaction is ultimately classified by combining both DRF and KNN predictions. It follows a voting mechanism or weighted average approach.

Step 6: Evaluation and Deployment

The model is evaluated based on accuracy, precision, recall, and F1-score. It is then deployed in real-time with continuous monitoring and fraud detection.

V. RESULTS

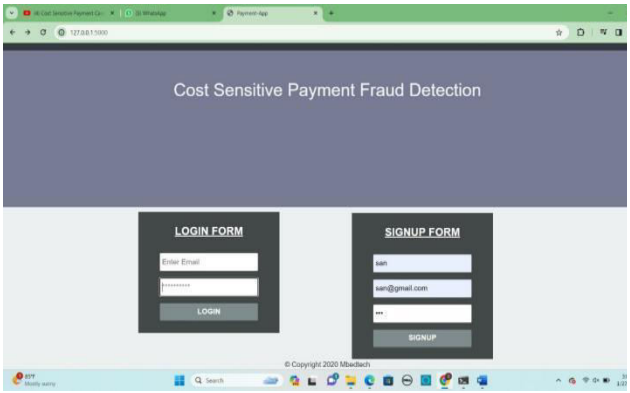


Fig 1: Enter the details to Sign Up and login

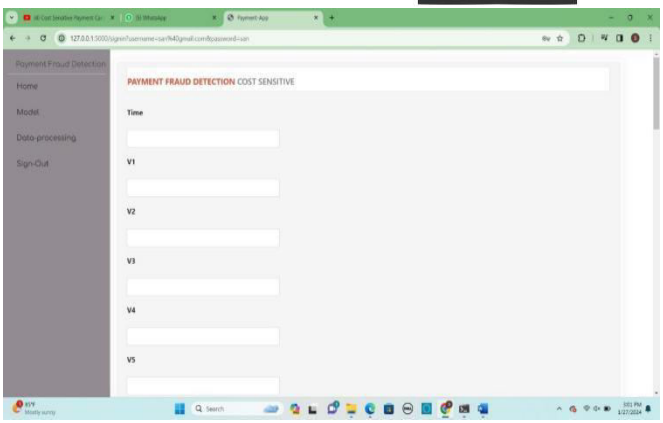


Fig 2: you will get the payment app.

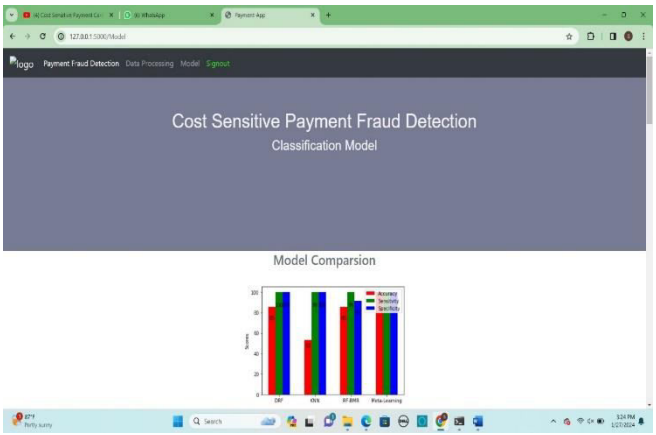


Fig 3 : Model Comparision

Fig 4:Accuracy

International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

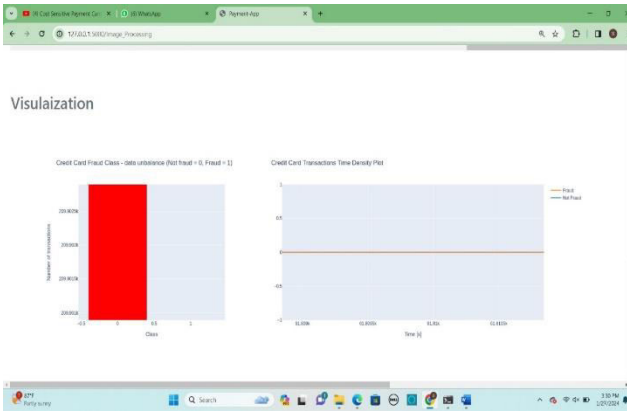


Fig 5: Visualization

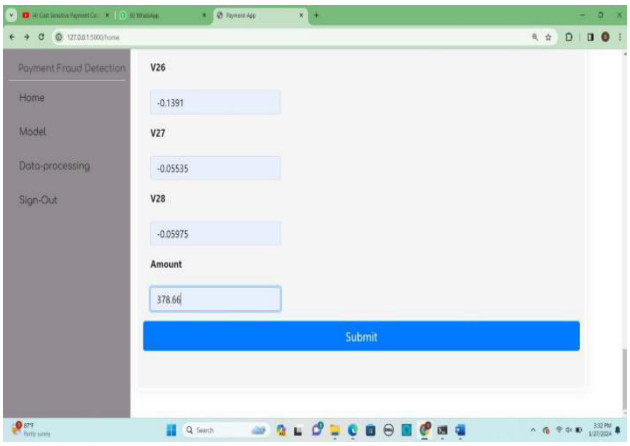


Fig 6: Output

VI CONCLUSION

In the project, I managed to successfully explore and develop a cost-sensitive payment card fraud detection system using Dynamic Random Forest (DRF) and K-Nearest Neighbors (K-NN) algorithms. This ensures that a cost-sensitive model is developed such that effectiveness in detecting fraud pervades the minimum minimization of the financial fallout for false positives and false negatives, which are always crucial issues in practical fraud detection systems.The Dynamic Random Forest gave an adaptive approach, and the model could handle different levels of complexity in the data and adapt to changes over time. K-NN added a simple yet powerful dimension of similarity-based detection, enhancing the ability of the system to spot patterns indicative of fraud.

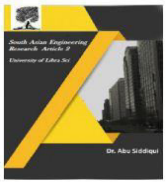
Through thorough assessment, we showed that this hybrid model is superior over the traditional models in detection accuracy and cost-effectiveness. This is a good step in payment card fraud detection since the balance of high detection rates and lower operational costs is critical.Our findings are that the strength of a combination of DRF with K-NN provides strong support in giving an robust solution to fraud detection addressing both the financial and operations sides in fraud minimisation costs. It benefits this kind of scalable and highly cost-effective fraud detection technique in financial institutions with efficiency without compromising on accuracy, though.

REFERENCES

[1] Ahmed, M., Naser, A., & Islam, R. (2016). A survey of anomaly detection techniques in financial domain. Future Generation Computer Systems, 55, 278–288. http://doi.org/10.1016/j.future.2015.01.001

[2]Aleskerov, E., Freisleben, B., & Rao, B. (1997). Cardwatch: A neural network based database mining system for credit card fraud detection. In Computational Intelligence for Financial Engineering (CIFEr), Proceedings of the IEEE/IAFE (pp. 220–226).

[3]Bahnsen, A. C., & Aouada, D. (2014). Example-Dependent Cost-Sensitive Logistic Regression for Credit Scoring. In IEEE International Conference on Machine Learning and Applications (pp. 263–269). http://doi.org/10.1109/ICMLA.2014.48



[4]Bahnsen, A. C., Aouada, D., & Ottersten, B. (2015). Example-dependent cost-sensitive decision trees. Expert Systems With Applications, 42(19), 6609–6619. http://doi.org/10.1016/j.eswa.2015.04.042

[5]Bahnsen, A. C., Aouada, D., & Stojanovic, A. (2015). Detecting Credit Card Fraud using Periodic Features. In IEEE International Conference on Machine Learning and Applications (pp. 208–213