International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Detect Suspicious Transactions and Identify Fraud Transactions in Banking Data Using Machine Learning

Mrs. Geetanjali Sharma[1], Dr. Shashi Bhushan[2], Dr. Ram Joshi[3], Dr. Asmita Manna[4], Mrs. Madhuri Amol Suryavanshi[5]

Submitted: 08/01/2024 **Revised**: 14/02/2024 **Accepted**: 22/02/2024

Abstract: The identification of potentially malicious transactions is a crucial and difficult activity for the purpose of preventing cyberattacks, which may have devastating effects on a company's finances, reputation, and legal standing. This is an extremely important problem for assuring the safety of investors in investment applications. When it comes to our scenario, the goal of identifying potentially suspicious transactions is to determine the typical patterns of behavior shown by application users and to pinpoint any circumstances that deviate significantly from these norms at the appropriate moment. Changing your password, logging in from several locations, and engaging in high-volume cash flow or trading activities are all examples of events that have a high level of significance for investing applications. It is very essential to promptly identify suspicious transactions and activity and to take the appropriate steps in order to protect the users as well as the assets and reputation of the firm. The proposed method for predicting fraudulent transactions begins with the dataset, which is pre-processed in the initial phase and then passed into further steps where, test/train split was performed, dataset spliced model parameters were defined, then genetic algorithm for feature selection was initiated, which returns weights for particular features after that model is loaded and those weights are passed into it, and model training was then initiated, and once training was completed, model M was returned.

***Keywords:** Suspicious Transactions, KNN , SVM , Random Forest , Adaboost ,Navies Bayes.*

1. Introduction

The terms "cyber" and "physical" are sometimes used interchangeably in the context of Cyber-Physical Systems (CPS), which are defined as "systems that include networking and controlling components that are diverse, federated, geographically distributed, and of a massive scale" [1, 2]. Wireless networking components, legacy parts, predictable network traffic, complicated network needs, and several control loops are all present in such systems. With the help of actuators and sensors, the system combines the physical and digital worlds (i.e., commodity servers and network constituents). Long-term and short-term assaults are both types of threats to CPS. Because to the intruder's interference, the system may fail completely [3].

Now more than ever, technical calculations and physical systems work together. Connected and pervasive systems (CPS) are comprised of a wide range of devices that are able to exchange data with one another and interact with the real environment (with high dependability) [4, 5]. As a whole, they work together with an eye toward both communication and computation (with monitoring and controlling procedures). Communicating and computing, controlling, and manipulating and monitoring are the three pillars of CPS. In order to ensure that all frameworks, including control centers and lower-level components, function together, the CPS uses a centralized coordination system (which utilize the wireless channel for communication).

To commit fraud against a monetary system for the purpose of gaining illicit financial gain by any group or person is considered to be an abuse of the system. With the emphasis on competitiveness in today's world, spoofing and other forms of deception may have devastating effects on a company. This is now a really significant problem. Credit card and bank account scams have been on the rise recently, costing banking and financial institutions billions of dollars annually. The BCPS is an important part of the modern economic landscape. Places where almost everyone is required to engage with it, either virtually or in person. Because of BCPS, business and public sector efficiency, profitability, and productivity have skyrocketed. Credit cards are now widely used for commercial transactions. There is a terrifying potential for new forms of fraud and assaults to be launched against such infrastructures. As one of the most integral parts of

[1]Research Scholar, JJTU, Rajasthan. Working as Assistant Professor, Computer Engineering Department, Pimpri Chinchwad College of Engineering, Nigdi,Pune, India
[2,]Senior Lecturer, Department of Computer & Information Sciences (CIS), Universiti Teknologi PETRONAS , Malaysia
[3]Dean Academics, Department of Information Technology, RSCOE,Pune, India
[4]Assistant Professor,Computer Engineering Department,Pimpri Chinchwad College of Engineering,Nigdi,Pune, India
[5]Assistant Professor,Computer Engineering Department,Pimpri Chinchwad College of Engineering,Nigdi,Pune, India
Email ID : [1]geetu2k3@gmail.com , [2]tyagi_shashi@yahoo.com [3]ramjoshi.comp@ gmail.com [4]asmita.manna@pccoepune.org, [5]madhurisuryavanshi.1122@gmail.com
** Corresponding Author: Geetanjali Sharma*
E-mail Id: geetu2k3@gmail.com

most people's day-to-day lives, the banking industry makes suspicious transaction detection in BCPS a pressing concern. A more reliable, efficient, and rapid strategy to detecting suspicious transactions is required in light of rising security concerns, making it imperative to research and implement effective AI technologies in the BCPS. BCPS needs suspicious transaction detection to boost security and trustworthiness for check and fraud prevention. Previous research on fraud detection has relied on tried and true methods. Consequently, work is needed on such systems that can learn from previously identified frauds and become adaptive in their detection of fraud in the future by using quantum mechanics.

In this paper key contributions below.

Existing work :

- Xgboost using boosting of trees , number of trees favours that class which returned by resulted class.
- Genetic algorithm flows selection first , after crossover ,mutation and fitness.

Proposed work :

- Xgboost used as baseline model, weights of error transfer from each trees uses initialized parameters specific for use case.
- Genetic algorithm selects weights of given parameters, exit to give best values of paramters.
- Adjusted weights for tree based models to predict frauds in financial payments
- Deduce behaviour of skewed financial data into normal on modelling
- Hypothesis based Feature Selection which reduces false positives
- Returns rules as well for rule based modelling to mitigate fraud at initial level fraudulent transactions.

In all, there are 5 chapters in this piece of writing. The next chapter, chapter 2, discusses related studies and theoretical elements of anomaly identification in financial data as well as applicable ML algorithms. This chapter follows the chapter that introduces the topic. The research methodology is presented in Chapter 3, along with descriptions of some of the chosen methodologies and the framework. The procedures of implementation as well as the empirical data are presented in chapter 4, and the conclusions are discussed in chapter 5.

2. Literature Review

Machine learning is a subfield of both artificial intelligence (AI) and computer science that focuses on simulating the way in which humans learn via the use of data and algorithms in order to achieve an ever-increasing degree of realism in the simulations created by these methods. Deep learning is another name for artificial intelligence, generally known as AI.

One of the most well-known and commonly used machine learning algorithms is called logistic regression. This technique belongs to the larger category of supervised learning, which is a learning approach. It is possible to make a prediction about the categorical dependent variable by employing a set of independent variables that have already been identified. Logical regression may be used to make predictions about the results of an experiment using a categorical dependent variable. As a direct consequence of this, the value that is returned must be either categorical or discrete. It might be yes or no, zero or one, the truth or a lie; yet, rather than presenting the exact value as 0 or 1, it displays the probability values that fall between the two numbers. This could mean that it could be either the truth or a falsehood. [6]

A technique for supervised machine learning that may be applied to problems requiring classification and regression is known as the Support Vector Machine, or SVM for short. Simply using this method's acronym as a stand-alone reference is typical practice. On the other hand, the most common use for it is in relation to concerns of classification. The support vector machine (SVM) approach requires that each data item be plotted as a point in an n-dimensional space (where n is the number of characteristics that are available). The value of a given location serves as a representation for the value of each individual feature. The value n indicates the total number of characteristics that you have. [7]

Introduction The information is continuously split in Decision Trees, which belong to the subfield of machine learning known as supervised learning. This is done in accordance with a parameter that the user specifies (which means that you explain what the input is and what the associated output is in the training data). [8]

The Random Forest technique, which is commonly regarded as one of the most well-known machine learning algorithms, is classified as an example of supervised learning, which is a more general classification. Classification and regression are two examples of types of activities that fall under the umbrella of machine learning and might possibly benefit from their utilization. It is based on ensemble learning, which is a strategy that integrates a number of distinct classifiers in order to solve a difficult issue and enhance the functioning of the model. This allows for more accurate predictions to be made. [9]

The k-nearest neighbour's technique is a non-parametric, supervised learning classifier that makes use of proximity to create classifications or predictions about an individual data point grouping. The approach was named after the k people who live the closest to a given data point. The technique was given its name after the k individuals who lived the shortest distances from one another. In certain communities, you could hear it referred to as KNN or k-NN instead. The most common use for this method is as a classification strategy due to the fact that it is predicated on the concept that points with similar features might be discovered in close proximity to one another. In spite of the fact that it may be used to problems requiring regression or classification, the notion behind it causes it to be utilized most often as a strategy for classifying things. [10]

The Naive Bayes algorithm offers a method for tackling classification issues that makes use of supervised learning as its primary data collection method. This methodology relies heavily on Bayes' theorem as its underpinning. The most important use for this technology is text categorization, which often requires the utilization of a high-dimensional training dataset. The Naive Bayes Classifier is an example of a classification algorithm that is not only straightforward but also incredibly efficient. It helps contribute to the creation of rapid machine learning models that are able to deliver predictions in a timely way. [Citation needed] It is a probabilistic classifier, which means that it based its predictions on the chance that a certain item will be recognized. This implies that it can accurately forecast the presence or absence of items. As a direct consequence of this, its forecasts could turn out to be rather accurate. [11]

Deep neural networks are a kind of artificial neural network (ANN) that include a significant number of hidden layers in between the network's input and output layers (DNN). DNNs, which are very similar to shallow ANNs and have the ability to express sophisticated non-linear connections, are becoming more popular.

In order to address problems that arise in the real world, such as classification, the primary function of a neural network is to take in a series of data, which are referred to as inputs, process those inputs using more complex algorithms, and provide results that can be used in the process of finding a solution to the problem. At this point in time, feed-forward neural networks are all that we are interested with. [12]

This research makes use of the oversampling method in addition to other tactics in order to get over the limitations put on the inquiry by the class imbalance problem. The findings that are generated by the models are analyzed by a variety of various kinds of assessors, and the results are then utilized to determine how successful the models are. The C4.5 decision tree and the Random Forest approach, when applied to the oversampled feature set, produce the best outputs in terms of accuracy and values for the Receiver Operating Characteristic (ROC). However, out of the three strategies, the Random Forest technique gives the best accurate forecasts. Nevertheless, the values of accuracy, recall, and F-measure might take on a number of distinct appearances depending on the context. [13] The research that is going to be done in the future is going to focus on finding the underlying causes of observed irregularities and developing quantitative metrics that can be used to anticipate the status of a computer system. These are going to be the topics of attention for the work that is going to be done.

Throughout the course of our investigation, we made use of a Multiple Classifiers System (MCS) on the two sets of data that are detailed below: credit card frauds (CCF) and (ii) credit card default payments (CCDP). In order to carry out precise anomaly detection, the MCS makes use of a method known as sequential decision combination. The following are some of the advantages of using this method: Our empirical investigation has shown that the MCS performs better than the previous research, particularly when recognizing irregularities that are uncommon occurrences in these two credit card data sets [14]. This is particularly the case when recognizing irregularities that are rare occurrences in the data. Our research aimed to demonstrate that the MCS is more effective than what had been found in earlier studies.

The CoDetect is able to concurrently recognize the feature patterns associated with fraudulent behaviors and operations in the financial sector. Some examples of these feature patterns are account balances and history of transactions. The proposed architecture has been shown, via exhaustive testing on both synthetic data and data acquired from the actual world, to be useful in the battle against financial fraud, notably money laundering [15]. The testing was carried out to provide evidence that the design is effective and productive in preventing financial fraud.

The model is educated and verified by making use of real-world corporate data, both financial and non-financial, that is publicly available on the internet. This data may be found in a variety of formats, including spreadsheets, databases, and charts. It has been shown to provide improved variable selection in addition to credit anomaly prediction with increased accuracy and resilience in comparison to other models that are considered to be state-of-the-art models. The findings not only validate that the incorporation of non-financial variables significantly improves the accuracy of the model's prediction of corporate credit anomalies, but they also demonstrate that critical financial factors are connected with identifying corporate credit anomalies. This is because the findings validate that the incorporation of non-financial variables significantly improves the accuracy of the model's prediction of corporate credit anomalies [16] These results show that there are important financial aspects that are related with the discovery of abnormalities in corporate credit, and they illustrate this connection.

The method was examined using two independent datasets: one with simulated traffic provided by the data communication and networking research group Orion at the state university of Londrina CICDDoS2019, a well-known anomaly detection dataset. The other dataset was internet-collected real-world data. Both sets of data came from Londrina's Computer Science Department. Both sets of data come from the State University of Londrina's Department of Computer Science. The results suggest using GRU networks and fuzzy logic to identify abnormalities in SDN and other applications. Research shows this. This method's success proves it. Several deep learning methods [17] were used to assess the system.

3. Proposed Work

3.1 Proposed flow architecture:

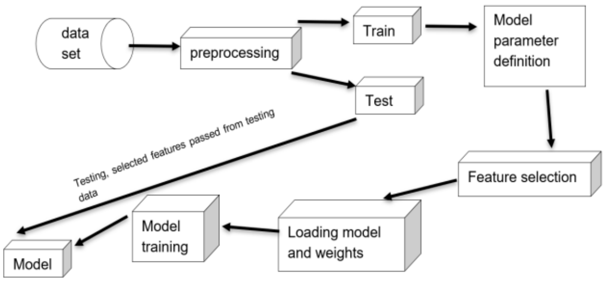
data set
preprocessing
Train
Test
Model parameter definition
Feature selection
Loading model and weights
Model training
Model
Testing, selected features passed from testing data

Figure 1. Working flow of the proposed methodology.

Figure 1 explains brief working flow of the proposed methodology to predict fraudulent transactions, flow begins with dataset which pre-processed in initial phase and passed into further steps where, test/train split performed , dataset spliced model parameters were defined then genetic algorithm for feature selection initiated which returns weights for particular features after that model is loaded and those weights are passed into it, and model training initiated once training is completed model M is returned by the methodology which further consumed by testing dataset to performing testing analysis.

3.2 Algorithm 1 : Feature Selection based on Hypothesis

Input : Dataset

Output : Preprocessed Dataset

```
Begin
 Load dataset into D
For each Column C in D dataset:
Calculate correlation crr between all C
If crr >0.6:
Drop C
Else if crr < 0.6:
Determine statistical test and generate Ci having better information gain.
If chai square test postive:
 Encode C
Else
Drop C
      Update all C in D
Return  D
```

This algorithm modifies the dataset and perform Feature Selection based on statistical properties of features, and returns the updated dataset. Algorithm begins with loading dataset D then iterate over columns of the dataset D, and check pearson correlation between pairs of the columns, and drooped column with pairs having it greater than 60 %, and also compepet descriptive statistics of all columns and engineered new columns. Correlation between categorical columns if negative then drop else encode columns, then dataset with updated columns returned by algorithm.

3.3 Algorithm 2 : Predicting fraudulent claims

Input : Dataset with preprocessed columns

Output : predict classes

```
Begin
Load dataset D
Split D into train and test
Initialize paralits = [max depth, scaling ratio, early Pruning,  column subsample for tree ]
Load XGboost model M
Apply genetic algorithm selecting best values for paralits
Store weights into W,
Assign W  trees
Set paralits into M
Fit M with train,W, paralits
Returns M
```

Algorithm returns model M trained over particular set of records or dataset through M variable, It begins with loading dataset and splitting into train and test. Then algorithm Initialize the parameters for modelling, model is loaded after parameters are defined then selecting best optimal values for parameters using genetic algorithm for modelling, and weights assigned to each parameter while mutation is performed in genetic algorithm is stored in W, Then Model fitted using all parameters, weights and training data, once model is trained internal validation is performed over model while training, then algorithm returns model M.

4. Implementation and Result

4.1 Dataset

Link : https://www.kaggle.com/datasets/ealaxi/paysim1

Dataset is synthetically generated by collecting data records from various financial institutions , and contains 11 columns , research in financial datasets crucial that 's why dataset collected from various source of transaction points.

PaySim as an method to such a problem. PaySim is aggregation of data records from the various private dataset to produce a synthetic dataset that resembles the normalized operation of payment transactions and injects malicious behaviour to later evaluate the performance of fraud detection methods.

4.2 Analysis Graphs

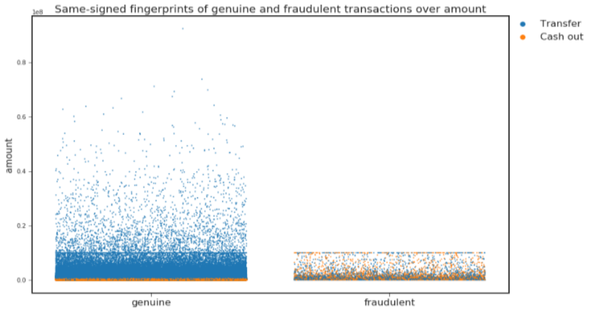

Fig 2. Shows, signed fingerprints of genuine and fraud transactions ratios.

Figure 2 shows, signed fingerprints of genuine and fraud transactions ratios, in which y-axis represents amount and x – axis shows type of case whether it is genuine or fraud cases, data points are hued based on type of transaction whether is transfer of payment or cash out , recorded into payments method. Based on figure it is observable that genuine vases have transfer type transaction more than cash out and also having high amount cases.

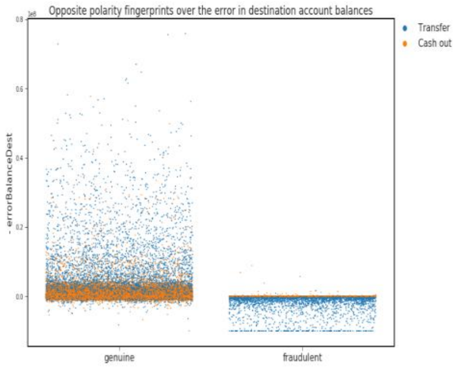

Fig 3. Shows records of error in destination account balances

Figure 3 shows records of error in destination account balances, in which y-axis gives error balance destination amount , and x-axis gives whether it is genuine payment or belongs to fraud cases these all information's are hued on mode of payment whether it is account transfer or cash out transaction.

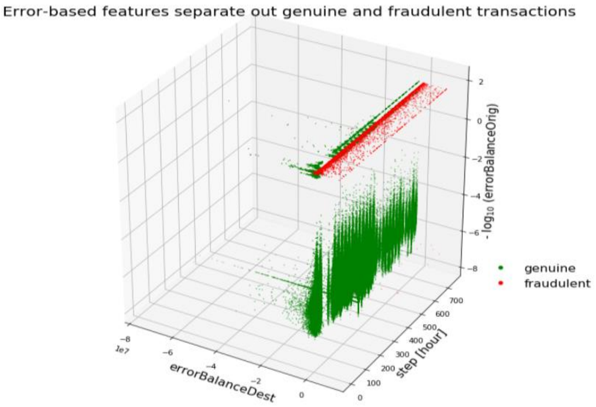

Fig 4. Represents 3D graph which separate out transactions based on error balance origin against error balance destination on genuine and fraud cases.

Figure 4 represents 3D graph which separate out transactions based on error balance origin against error balance destination on genuine and fraud cases .it gives pattern of transactions in particular spam of time and all these parameters are hued on genuine and fraud cases.

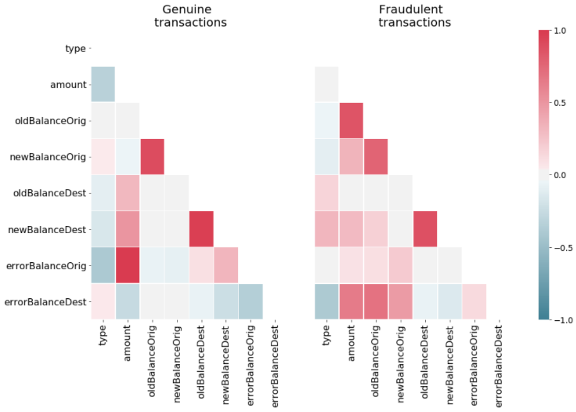

Fig 5. Represents correlation between essential features of the dataset with respect to genuine and fraud cases.

Figure 5 represents correlation between essential features of the dataset with respect to genuine and fraud cases which gives whole some idea for features selection part, where more the bluish higher the negative correlation represented between particular columns, and higher the positive correlation if more red colour is in relation of particular features.

4.3 Results

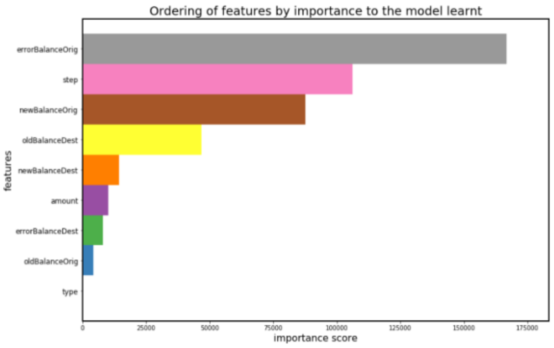

Fig 6. Represents the feature importance from set of features used for model training top bar represented features have more contribution.

Figure 6 represents the feature importance from set of features used for model training top bar represented features have more contribution, in figure y-axis represented features names and x-axis repented importance score, from the importance chart it is observable that error balance origin plays vital role into decision of whether that particular transaction is genuine or fraud .

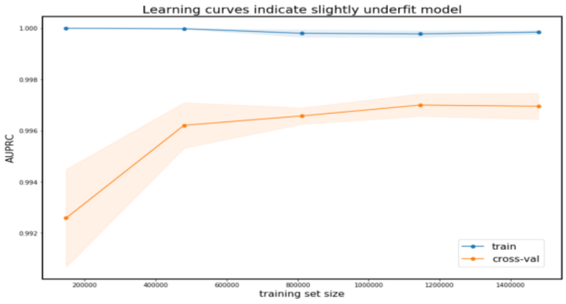

Fig 7. Represents training and validation scores of model.

Figures 7 represents training and validation scores of model, in which y-axis repents AUPRC and x-axis represents training size it observable from the graph that as heading towards completion of training part models stabilizes it results, reaches to global scores which shows proper balance between bias and variance of model.

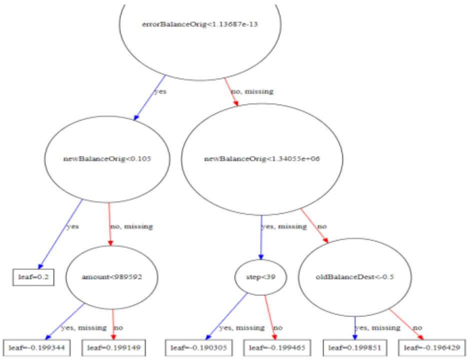

Fig 8. Generic rule from surrogated tree.

Figure 8 gives generic rule from surrogated tree where top open node is root for that particular transaction records blue lines for positive condition and red lines for negative cases, and records decisions where taken based on the flow of values in the surrogated rule tree.

Table 1. Experiments performed various financial institutions with different existing solutions table contains comparative studies of, those experiments with proposed model.

Models	**AUC**	**Recall**	**Precision**	**F1 score**
KNN [18]	0.92	0.61	0.30	0.28
SVM [19]	0.90	0.63	0.29	0.30
Random Forest [20]	0.89	0.70	0.48	0.42
Adaboost[21]	0.92	0.75	0.62	0.59
Navies Bayes [22]	0.95	0.81	0.74	0.78
Proposed	0.98	0.89	0.91	0.90

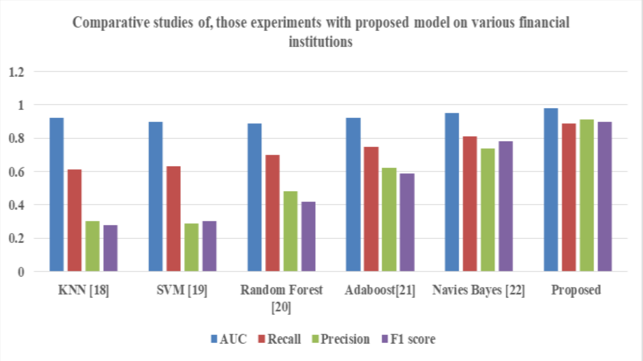

Table 9. Experiments performed various financial institutions with different existing solutions table contains comparative studies of, those experiments with proposed model.

Table 1 and figure 9 shows several experiments performed over same dataset with different existing solutions table contains comparative studies of, those experiments with proposed model , it is observed from the table that results moving towards higher sides as models starting getting complexes as, even though probalistic model such as navies Bayes performed better than, tree based solutions, but introduction of feature selection based on genetic selection makes tree based models to performed better than existing solutions , and gives best F1 score for detection of frauds in financial payments.

4.4 Results on Dataset

Data set link : https://www.kaggle.com/code/turkayavci/fraud-detection-on-bank-payments/data

Dataset Description: banking fraudulent behavior increasing day by day as digital transactions are increasing so, analyzing and capturing more precisely predicting becoming important part in the financial institutions, which is derived from the PWC survey 2018 on banking organizations. Dataset taken from kaggle which contains few finding from PWC survey and data transaction records, which used to train and test of proposed model.

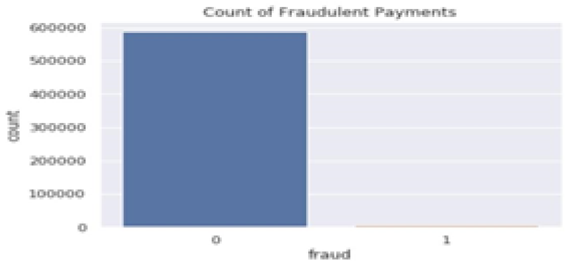

Fig 10. Represents data ratio with fraud and count of total number of cases.

Figure 10 represents data ratio between fraud to normal cases , in blue color it highlighted normal cases and in orange color it shows fraudulent cases represent via 1 on fraud or X-axis., and y- axis it shows count of cases .

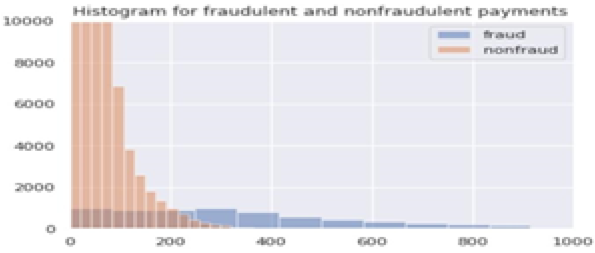

Fig 11. Represents distribution of the classes presents in dataset.

Distribution presented in figure 11 shows that, non fraudulent cases and fraudulent both records are left skewed not follows normal distribution which signifies that, in order to retrieve best results from the model normalization over dataset is needs to be performed so, that results on data which not skwed towards any end model performs equally where it skwed.

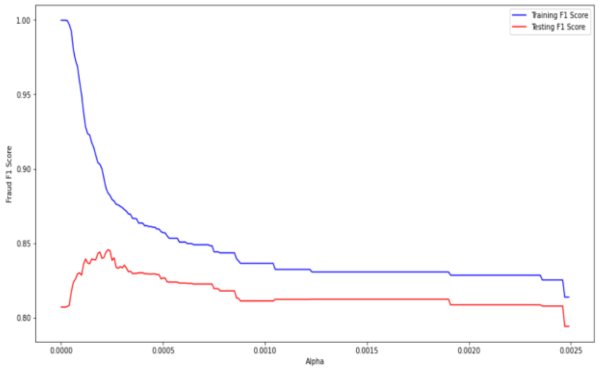

Fig 12. Represents alpha relationship with f1 score.

Above figure 12 shows training and validation performance of the tree based models on different alpha values, it was observed from the curve that regularization works, as value of increasing as distance between training and validation lines reducing and seem to meet on higher values of alpha. Which shows over fitting handled by higher values of alpha.

Table 2. Experiments performed over banking fraudulent behavior with different existing solutions table contains comparative studies of, those experiments with proposed model.

Models	AUC	Recall	Precision	F1 score
KNN [18]	0.89	0.67	0.32	0.29
SVM [19]	0.76	0.65	0.27	0.37
Random Forest [20]	0.80	0.75	0.47	0.47
Adaboost [21]	0.91	0.74	0.66	0.55
Navies Bayes [22]	0.91	0.88	0.77	0.79
Proposed	0.92	0.90	0.95	0.97

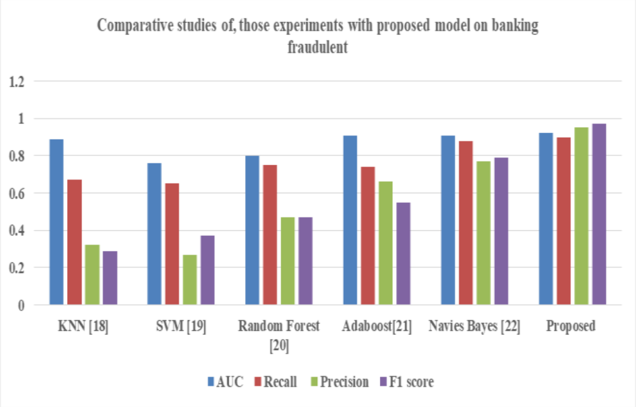

Fig 13. Experiments performed over banking fraudulent behavior with different existing solutions table contains comparative studies of, those experiments with proposed model.

Table 2 and figure 13 shows comparative analysis of various existing solutions with proposed one it was observed from the table that, results of the proposed model outperforms over the existing solutions, on an average improvement of 8% , which shows that model impact financial organization with proper margins as 9% detection in downfall supports proper management of resource which might be utilized to manage the risk properly.

5. Conclusion

A significant competitive advantage for many companies operating in a variety of markets would be the capacity to recognize fraudulent activity in transaction data. The majority of data sets can only be handled effectively with some kind of automation due to the large amounts of data they include. Because of the complexity of the data, it is difficult to build algorithms that are hard-coded and can predict every possible kind of anomaly. Because of these challenges, unsupervised machine learning is a natural starting point for attacking the problem of fraud detection, and Isolation Forest is a strong representative of the family of ML techniques that are useful for spotting outliers. Proposed model , it is observed from the table that results moving towards higher sides as models starting getting complexes as, even though probalistic model such as navies Bayes performed better than, tree based solutions, but introduction of feature selection based on genetic selection makes tree based models to performed better than existing solutions , and gives best F1 score for detection of frauds in financial payments.

Author contributions

Geetanjali Sharma: Conceptualization, Methodology, Software, Field study, Data curation, Writing-Original draft preparation, Software, Validation., Field study. **Dr. Shashi Bhushan, Dr. Ram Joshi, Dr Asmita Manna , and Mrs. Madhuri Amol Suryavanshi**: Visualization, Investigation, Writing-Reviewing and Editing.

Conflicts of interest

The authors declare no conflicts of interest.

References

[1] Al-Momani and S. A. Aljawarneh, "Fraudulent Transactions Prediction Using Deep Neural Network," 2022 International Conference on Engineering & MIS (ICEMIS), 2022, pp. 1-7, doi: 10.1109/ICEMIS56295.2022.9914349.

[2] M. Demirdag, E. Bay, G. Yuceturk, S. M. Yalçinkaya and I. U. Sayan, "Anomaly Detection in Investors' Activities and Transactions via Rule-Based and Advanced Technique," 2022 International Joint Conference on Information and Communication Engineering (JCICE), 2022, pp. 11-14, doi: 10.1109/JCICE56791.2022.00013.

[3] R. A. L. Torres and M. Ladeira, "A proposal for online analysis and identification of fraudulent financial transactions," 2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA), 2020, pp. 240-245, doi: 10.1109/ICMLA51294.2020.00047.

[4] Catherine, Denny and M. R. Shihab, "Bank Account Classification for Gambling Transactions," 2021 3rd East Indonesia Conference on Computer and Information Technology (EIConCIT), 2021, pp. 302-308, doi: 10.1109/EIConCIT50028.2021.9431874.

[5] T. Zhu et al., "Detecting and Warning Abnormal Transaction of Virtual Cryptocurrency Based on Privacy Protection Framework," 2022 IEEE 7th International Conference on Smart Cloud (SmartCloud), 2022, pp. 74-82, doi: 10.1109/SmartCloud55982.2022.00018.

[6] Sarker, IH Machine Learning: Algorithms, Real-World Applications and Research Directions. SN COMPUT. SCI. 2, 160 (2021). https://doi.org/10.1007/s42979-021-00592

[7] Cao, L., Tay, F. Financial Forecasting Using Support Vector Machines. Neural Comput & Applic 10, 184–192 (2001). https://doi.org/10.1007/s005210170010

[8] Jin, M., Wang, H., Zhang, Q. et al. Financial Management and Decision Based on Decision Tree Algorithm. Wireless Pers Commun 102, 2869–2884 (2018). https://doi.org/10.1007/s11277-018-5312-6

[9] Fawagreh, K., Gaber, M.M., Abdalla, M. (2022). Pruned Random Forests for Effective and Efficient Financial Data Analytics. In: Derindere Köseoğlu, S. (eds) Financial Data Analytics. Contributions to Finance and Accounting. Springer, Cham. https://doi.org/10.1007/978-3-030-83799-0_7

[10] Ban, T., Zhang, R., Pang, S., Sarrafzadeh, A., Inoue, D. (2013). Referential kNN Regression for Financial Time Series Forecasting. In: Lee, M., Hirose, A., Hou, ZG., Kil, R.M. (eds) Neural Information Processing. ICONIP 2013. Lecture Notes in Computer Science, vol 8226. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-42054-2_75

[11] Brumen, B., Rozman, I., Černezel, A. (2014). Observing a Naïve Bayes Classifier's Performance on Multiple Datasets. In: Manolopoulos, Y., Trajcevski, G., Kon-Popovska, M. (eds) Advances in Databases and Information Systems. ADBIS 2014. Lecture Notes in Computer Science, vol 8716. Springer, Cham. https://doi.org/10.1007/978-3-319-10933-6_20

[12] Yan, H., Ouyang, H. Financial Time Series Prediction Based on Deep Learning. Wireless Pers Commun 102, 683–700 (2018). https://doi.org/10.1007/s11277-017-5086-2

[13] P. A. Samarakoon and D. A. S. Athukorala, "System abnormality detection in stock market complex trading systems using machine learning techniques," 2017 National Information Technology Conference (NITC), 2017, pp. 125-130, doi: 10.1109/NITC.2017.8285660.

[14] S. N. Kalid, K. -H. Ng, G. -K. Tong and K. -C. Khor, "A Multiple Classifiers System for Anomaly Detection in Credit Card Data With Unbalanced and Overlapped Classes," in IEEE Access, vol. 8, pp. 28210-28221, 2020, doi: 10.1109/ACCESS.2020.2972009.

[15] D. Huang, D. Mu, L. Yang and X. Cai, "CoDetect: Financial Fraud Detection With Anomaly Feature Detection," in IEEE Access, vol. 6, pp. 19161-19174, 2018, doi: 10.1109/ACCESS.2018.2816564.

[16] S. Han, "Semi-Supervised Learning Classification Based on Generalized Additive Logistic Regression for Corporate Credit Anomaly Detection," in IEEE Access, vol. 8, pp. 199060-199069, 2020, doi: 10.1109/ACCESS.2020.3035128.

[17] D. M. Brandão Lent, M. P. Novaes, L. F. Carvalho, J. Lloret, J. J. P. C. Rodrigues and M. L. Proença, "A Gated Recurrent Unit Deep Learning Model to Detect and Mitigate Distributed Denial of Service and Portscan Attacks," in IEEE Access, vol. 10, pp. 73229-73242, 2022, doi: 10.1109/ACCESS.2022.3190008.

[18] D Rzayeva, S Malekzadeh , Combination of Deep Neural Networks and K-Nearest Neighbors for Credit Card Fraud Detection - arXiv preprint arXiv:2205.15300, 2022 - arxiv.org

[19] Zhang, D. , Bhandari, B. and Black, D. (2020) Credit Card Fraud Detection Using Weighted Support Vector Machine. Applied Mathematics, 11, 1275-1291. doi: 10.4236/am.2020.1112087.

[20] Lin, T.-H.; Jiang, J.-R. Credit Card Fraud Detection with Autoencoder and Probabilistic Random Forest. Mathematics 2021, 9, 2683. https://doi.org/10.3390/math9212683

[21] Randhawa, K.; Loo, C.K.; Seera, M.; Lim, C.P.; Nandi, A.K. Credit Card Fraud Detection Using AdaBoost and Majority Voting. IEEE Access 2018, 6, 14277–14284

[22] Kaur, B.J., Kumar, R. (2020). A Hybrid Approach for Credit Card Fraud Detection Using Naive Bayes and Voting Classifier. In: Pandian, A., Palanisamy, R., Ntalianis, K. (eds) Proceeding of the International Conference on Computer Networks, Big Data and IoT (ICCBI - 2019). ICCBI 2019