International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 12 | Dec 2020 www.irjet.net p-ISSN: 2395-0072

# Credit Card Fraudulence Transactions Identification with the Aid of Machine Learning Methodologies

**Vaishnavi Mahalle[1], Yashashri Rathi[2], Shivani Pote[3], Khushboo Shriwas[4], Prof. Kalyani Gholap[5]**

*[1,2,3,4]Student's Dept. of Computer Science & Engineering, Dhamangaon Education Society's College of Engineering and Technology, Dhamangaon Rly., Maharashtra, India.*
*[5]Assistant Professor, Dept. of Computer Science & Engineering, Dhamangaon Education Society's College of Engineering and Technology, Dhamangaon Rly., Maharashtra, India.*

---

**Abstract** - *In these days, due to technology enhancement the credit card became a very common and popular item instead of carrying physical currency. It helps in providing cashless shopping over the globe. Extortion function happens only during on the web installment as Master Card number is adequate to make an exchange which will be on the Visa to make online payment however for disconnected installment secret phrase will be asked souring disconnected exchange fakes can't happen. The proposed framework's main necessity of identification of fraudulent transaction of all the transactions made through credit cards. The proposed framework is mentioned implemented through various popular machine learning algorithms such as KNN, logistic regression, SVM, decision tree, and random forest. The proposed framework was implemented on the famous dataset which was freely available on Kaggle. Overall, we got 100 % accuracy after using Lasso feature selection with various machine algorithms*.

***Key Words***: **Charge card, segregated backwoods, Local anomaly factor, Fraud identification, Data mining.**

## 1. INTRODUCTION

Due to a quick headway in electrical business innovation, the utilization of Visas has substantially expanded. As cost card turns into probably the most mainstream technique of installment for both online-only as regular buy, instances of misrepresentation connected to it are the same rising. A few of many ways that money can easily be stolen from the Visa are Dumpster, Skimming, Pharming, and Phishing driving. Programmers and fraudsters are becoming adept and complex more at controlling net convention, net dialects as well as instruments to or perhaps find any shortcoming that they can utilize. Along these lines, the net exchange misrepresentation is multiple times higher compared to in-store extortion.

MasterCard based buys may be sorted into 2 kinds: 1) physical card and 2) virtual card. In a physical-card-based purchase, the cardholder presents the card of his reality to the dealer to make an installment. To do fake exchanges in that kind of procurement, an assailant should carry the charge card. If the cardholder does not realize the loss of card, it can encourage significant cash associated misfortune to the MasterCard business. In the 2nd kind of procurement, simply a few substantial info about a (card number, termination date, and secure code) are necessary to create the installment. Such buys are ordinarily accomplished on the web or perhaps via telephone. To submit misrepresentation in these sorts of buys, a fraudster simply has to understand the card subtleties. In many cases, the veritable cardholder does not realize that another individual has seen or even taken the card data of his. The most effective way to perceive the kind of misrepresentation is breaking down the spending layouts on every card as well as to sort out any irregularity as for the "standard thing" spending models. Extortion location reliant on the probe of current acquisition info of cardholder is a promising strategy to reduce the speed of fruitful Visa cheats. Deviating from that kind of instance is a probable threat to the system.

Develop a Visa misrepresentation recognition version that could sufficiently identify cheats from an imbalanced dataset.

Fraud identification design thought each trait likewise without offering inclination to any kind of recognition in the dataset • Proposed misrepresentation recognition design creates sort lawful exchange layout (client purchasing individual conduct standard) as well as extortion exchange layout (fraudster private conduct standard) for each prospect and hence converted the imbalanced MasterCard exchange dataset into an equalization a person to deal with the problem of unevenness.

In recent years the fraudulent transactions based on credit cards is been increasing across the globe. This issue can be identified by considering the instances of Europe and the US which are considered to be economically strong countries. The evolution of fraudulent transactions in both instances can be represented as mentioned in the figure-1.

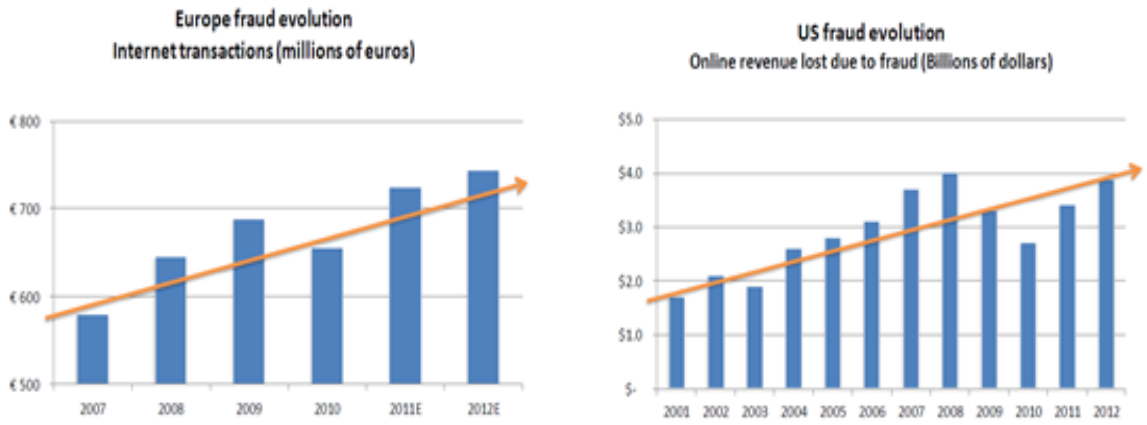


Figure 1: Impact of Credit Card Fraud Detection

The document is organized into various sections such as section-2 discusses the credit card fraud detection related aspects in the perception of a literature survey, section-3 discusses methodologies that are utilized to implement the proposed framework which includes the discussion of the dataset as well as the system requirements, section-4 discusses the results obtained through the implemented methodologies, and section-5 provide the conclusion obtained from the proposed as well as the implemented framework.

## 2. LITERATURE REVIEW

Wen-Fang YU et al 2009 [1] states that that in the current time of originality extraordinarily in the net trade as well as banking, the switches by the ace cards have been growing rapidly. The Mastercard comes to be the greatly utilizable hardware for net shopping. This particular expansion being utilized reason considerable damage and misrepresentation cases furthermore. The authors proposed a framework that describes the exception mining method that can perceive extortion cases much more than irregularity discovery. On the out chance that this specific evaluation is utilized on all of the identification frameworks, at that time we perceive the cheats rapidly utilize a little adversary of extortion methodologies to reduce the misfortunes earlier decline odds. Ounacer, S. et al 2018 [2] state that in today's moment of originality exceptionally in the net industry and banking, the switches by the ace cards have been growing rapidly. Strategic relapse, Decision tree, SVM, etc are a couple of ways to contend with recognizing abnormalities. Be that as it might, these methodologies are restricted because they're managed calculations that are ready by the marks to recognize climate the switches are misrepresentation or perhaps not. The authors proposed a framework that is usually to be determined by far the most noteworthy no of exchange with nearly all noteworthy accuracy, in that they also thought about a variety of solo methodologies for misrepresentation identification like one group SVM, K implies, LOF as well as Isolation Forest to find out the ideal method. IbatissiamBenchaji et al 2018 [3] state that in the current time of originality extraordinarily in the net trade as well as banking, the switches by the ace cards have been growing rapidly. The main goal of the proposed framework is upgrading the business presentation of the purchased instances in the imbalanced dataset for that they proposed the unaided assessment strategy reliant on the hereditary calculation as well as K implies grouping.

Andrea Dal Pozzolo et al 2018[4] portrays that in the current time of originality uniquely in the net trade as well as banking, the switches by the ace cards have been growing rapidly. The Mastercard come to be the exceptionally utilizable hardware for net shopping. This paper has three considerable goals such as the proposed development of the extortion location problem which depicts the problem for the exercise of misrepresentation identification framework that searching the exchanges day by day time with the assistance of the modern-day accomplice, they similarly program and get to learn the process which deals with the category unevenness, they learn the effect of concept float as well as category unbalance in a certifiable info stream that contains ¾ billion exchanges for more than three seasons. Lutao Zhang et al 2018[5] depict that with the expansion of the business that is online, exchanges are the same expanding in which several of them were extortion. To perceive the misrepresentation exchange, it's critical to extricate authentic exchange documents on the conduct profile of the clientele. To speaks to the BPs of the prospect the Markov chain design is recognized. Whose exchange practices are steady this will affect them. The Mastercard comes to be a greatly utilizable gear for net shopping. So in this particular paper, they proposed a constant chart of conduct profiles that speaks to the smart relations on the most out request based model. SP Maniraj et al 2019[6] things which in today's time of originality uncommonly in the net trade as well as banking, the switches by the ace cards has been growing rapidly. The MasterCard comes to be the greatly utilizable hardware for net shopping. It's a great deal essential to quit the misrepresentation exchanges since it affects our financial ailments after a few times the peculiarity identification is having a number of the significant programs to perceive the extortion find. Such problems may similarly be tackle with the assistance of Data science with the mix of AI. With

this particular technique, the authors have concentrated on pre handling informational collections analyzing notwithstanding the sending of several peculiarity location calculations as Isolation Forest calculation only as Local Outlier Factor on the PCA evolved Master Card Transaction data.

Sara Makki et al 2019[7] portray that in the current time of originality extraordinarily in the net industry and banking, the switches by the ace cards have been growing rapidly. The Mastercard comes to be the greatly utilizable hardware for net shopping. This expansion is utilized reason considerable damage and extortion cases furthermore. It's a great deal essential to quit the misrepresentation exchanges since it affects our money associated ailments after a few time the abnormality identification is having several significant programs to understand the extortion discovery. The paper fundamentally centered across the arrangement which discusses the irregularity problem of the order they check out the solution for misrepresentation finds utilizing AI computations. They besides learn the summed up outcomes and shortcomings they get making utilize of charge card misrepresentation called dataset. They give us the conclusion that the imbalanced order is ineffectual when the info is greatly imbalanced. With this article, they examined that the present strategies are outcomes of many phony alerts, which are costlier. Maes, Sam, et al. 2002 [8] proposed an automated framework for the identification of the fraudulent transaction of all the transactions made through credit cards. The authors also identified the prominence of the fraud transaction impact on the financial institutions. The proposed framework was implemented based on the artificial neural network as well s the Bayesian belief networks. The proposed framework has shown a significant result which can be utilized as motivation to improve the model further. Fu, Kang. et al. 2016 [9] identified the improvement in the usage of the credit card and at the same time the impact of fraudulent transactions. In that aspect, the authors proposed a fraudulent transaction identification system based on the convolutional neural network(CNN). CNN based system was developed by keeping the intuition that it can identify the inherent patterns in the transactions made through credit cards. Roy, A. et al. 2018 [10] detected various frauds and their impact on the financial institutions that are located in North America in the year of 2017. Also, detected the results provided by the fraud detection systems developed by using deep learning models able to attain comparable efficient results. So, the authors proposed a framework for the identification of fraudulent transactions based on one of the popular deep learning methodologies, LSTM. For this framework implementation, about 80 million instances of a dataset were utilized.

## 3. METHODOLOGY

The proposed methodologies are employed in this specific venture, for finally recognizing the fakes in the Visa framework. The examination is created for different AI calculations, for instance, Logistic Regression, Decision Trees, Random Forest, to discover which calculation provides suits the best and maybe modified with MasterCard sellers for knowing misrepresentation exchanges. The Figure given below displays the development graph for talking to the common framework structure.

### 3.1 Dataset and System Requirements

The dataset utilized was openly available credit card transaction-related to the identification of fraud detection during September 2013 in Europe in Kaggle. This data consists of 2, 84, 807 instances which are of having two classes such as fraudulent and genuine. Total instances of fraud are 492, and total instances of genuine are 2, 84,315. The distribution of these cases can be represented as mentioned in figure - 2. Various system requirements are utilized for the implementation of the proposed framework is mentioned in table - 1.

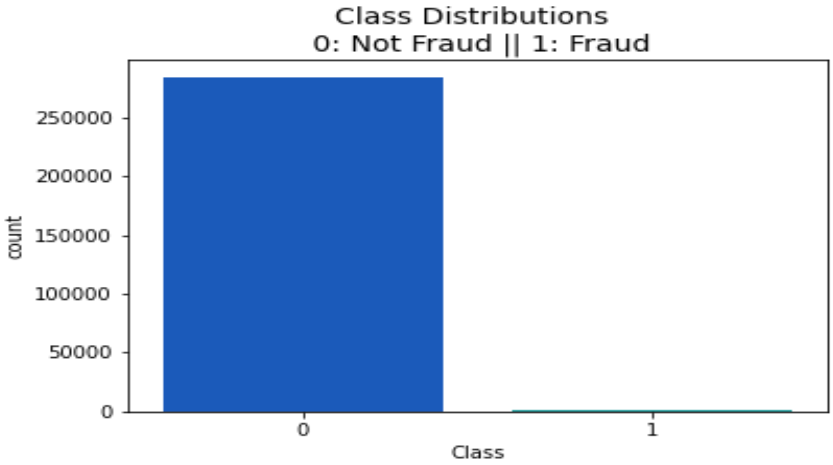


Figure 2: The Distribution of classes of the considered dataset

Table 1: System Requirements

| Hardware Requirements | Software Requirements |
|---|---|
| **RAM: Minimum 1 GB or above** | Operating System: Windows 7 or above |
| **Hard Disk: 500 MB or above** | Tool Utilized: ANACONDA |
| **Processor: I3 or above** | Navigator, Spyder, Online tool – Google Colab |

Various methodologies are utilized for the proposed framework are random forest, decision tree, support vector machine(SVM), logistic regression, and KNN. These methodologies are discussed as follows.

### 3.2 Random Forest Methodology

It is a particular design that is essentially an ensembling methodology based classification model, at this particular blending classification model utilizes and blends numerous decision tree-based classification models. The primary outline at the back of making utilization of numerous trees to have the ability for instructing the considered model of trees adequately which implies that the impact of each one of those is available in terms of a product. Once the modeled tree was generated, the result obtained is blended with the majority[11]. It utilizes numerous decision tree-based classification models to ensure the dependency of every one of those is actually on specific processing of the dataset, the same dissemination through the modeled decision trees. This specific design has the efficiency of assessing mistakes in a category public of distorted information datasets. It may be utilized to resolve mutual distinctions in addition to regression issues.

### 3.3 Decision Tree Methodology

It is one of the most commonly utilized analytical modeling methodologies. According to the title of the product, it's constructed in the type of a tree as a framework. This particular framework is perhaps utilized in the scenario of a multidimensional analysis in which there are several sessions present[12]. The preceding information also called the previous vector is utilized to produce a unit that may be utilized to calculate the worth of the result depending on the response of input being presented. There are various numerous nodes in each node and a modeled tree resembles various other vectors. Termination of the modeled tree at a leaf node in which the representation of every one of such nodes represent a probable effect or perhaps result.

### 3.4 Support Vector Machine(SVM) Methodology

A classification is solved by SVM and regression problems. This algorithm plots every instance of the sample as a stage in the multidimensional dataset. The value of a specific coordinate implies each value of the variable. Thereafter, this algorithm is going to detect the best hyper-plane is in a position to distinguish categories[13]. This algorithm is effective in attaining the best decision frontier between categories. Nevertheless, this algorithm doesn't operate effectively with datasets that have distorted category distribution, errors, and corresponding category samples. The algorithm related parameters can be changed to obtain classier much more resistance to errors as well as for working better for healthy datasets. But when it requires distorted

datasets, minority categorical samples might contemplate as errors. Thus, minority categorical samples will be disregarded completely by this algorithm.

### 3.5 Logistic Regression Methodology

It's essentially a statistical style and that makes utilize of a logistic feature to model a binary reliant adjustable. The particular design is primarily utilized with the possibility of the existence of a two-class classification problem. The effective working on categorical linear separability. The ratio of the odds is the intuition utilizing that can determine the logistic functionality[14]. This functionality is the likelihood of an outcome happening and it can be represented as mentioned in equation-1.

Odds Ratio = p / (l - p) (1)

The logit feature is the logarithm of the odds ratio. It takes input in the assortment of [0,1] and also transforms them to values over the real number range. The logit feature could be identified as mentioned in equation-2.

F(p) = log(p / 1-p) (2)

### 3.6 K-Nearest Neighbour(KNN) Methodology

This methodology develops the classic purpose through a popular vote of its neighborhood's neighboring information aspects. This illustrates the way KNN detects the category of a fresh sample. Assume the selection of neighbors, k D five, as well as the Euclidean metric which is a distance metric. This algorithm classier is going to and the closest samples to the fresh samples[15]. The distance metric among the target sample as well as the fresh sample in a dimensional-space is estimated utilizing the metric as-per (four), where p is two. Out of neighbors, the training samples of category A is greater than the training samples of category B. Consequently, the new sample is categorized as category A. This particular instance illustrates that this algorithm is expected to bias concerning the vast popularity of the category. The possibility of the fresh sample to be categorized as category B is pretty small in contrast to Class A. [16-21]

### 3.7 Evaluation Metrics

The proposed framework implemented using various machine learning algorithms here to evaluate the performance steps in the charge card fraud detection dataset. While predicting the class of the transaction made through credit card then the four scenarios arise such as true positive denoted by A, true negative denoted B, false-positive denoted by C, and false-negative denoted by D. then the resultant formulae for various evaluation metrics such as accuracy, precision, and specificity can be represented as mentioned in equations-3, 4, and 5.

Accuracy = (A+B) / (A+B+C+D) (3)

Precision = A / (A+C) (4)

Specificity = B / (B+C) (5)

### 3.8 Implemented Algorithm and Flow Chart

The proposed framework implemented using various methodologies as mentioned in the above section. The basic algorithm utilized as a backbone of the proposed framework as discussed as follows:

Step-1: Input the considered dataset.

Step-2: Partition the dataset into two sets such as training and testing datasets.

Step-3: Training dataset utilized for the phase of training to identify the various features from the considered training dataset.

Step-4: Predict the various classes such as fraud and genuine using the Testing dataset for the phase of testing

Step-5: Based on the prediction, evaluate the performance of the model using various metrics such as accuracy, precision, and specificity.

The flowchart of the algorithm can be represented as mentioned in the figure-

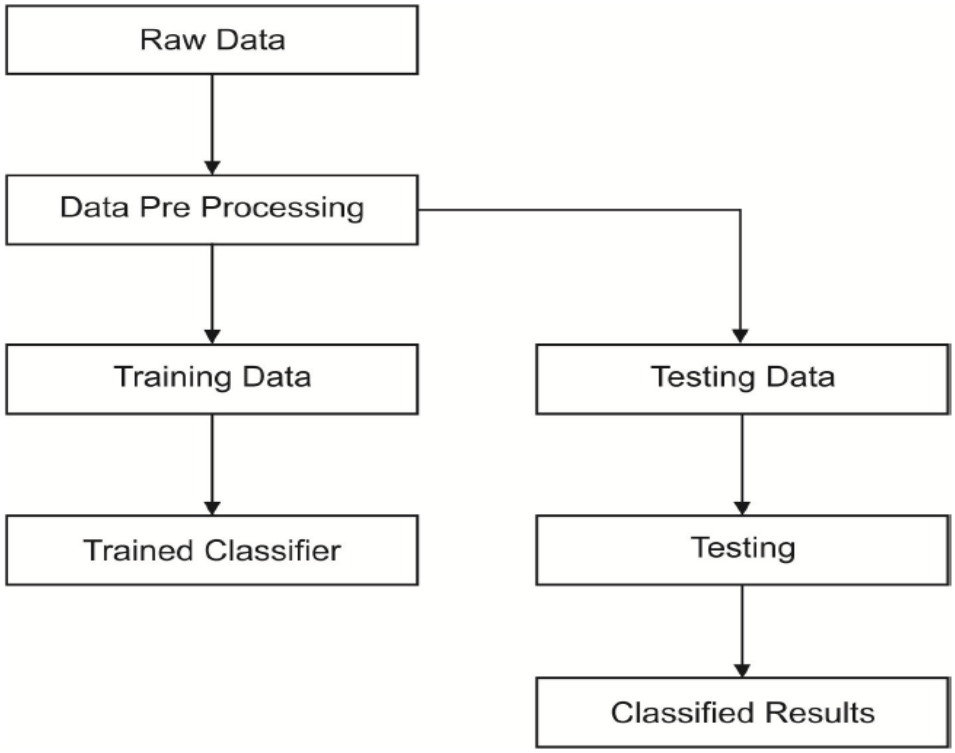


Figure 3: The flowchart of the proposed framework

**4. RESULT ANALYSIS**

The framework is to identify the fraudulent transactions of credit card based dataset considered in September 2013 in Europe. The performance of the framework is evaluated using evaluation metrics such as accuracy, precision, and specificity. There are four categories identified during the classification such as True Positive(TP) indicates the non-fraudulent transactions able to detect as non-fraudulent transactions, True Negative(TN) indicates the fraudulent transactions able to detect as fraudulent transactions, False Positive(FP) indicates the non-fraudulent transactions able to detect as fraudulent transactions, False Negative(FN) indicates the fraudulent transactions able to detect as non-fraudulent transactions.

The obtained results were mentioned in two forms such as confusion matrix and table of evaluation metrics. The evaluation metrics were evaluated for five different machine learning models implemented such as logistic regression, KNN, SVM, decision tree, the random forest along with the proposed algorithm represented as mentioned in table-2. One can observe that the results are highly satisfactory in the case of logistic regression, KNN, SVM, decision tree. The proposed framework able to attain better accuracy when compared to the random forest algorithm.

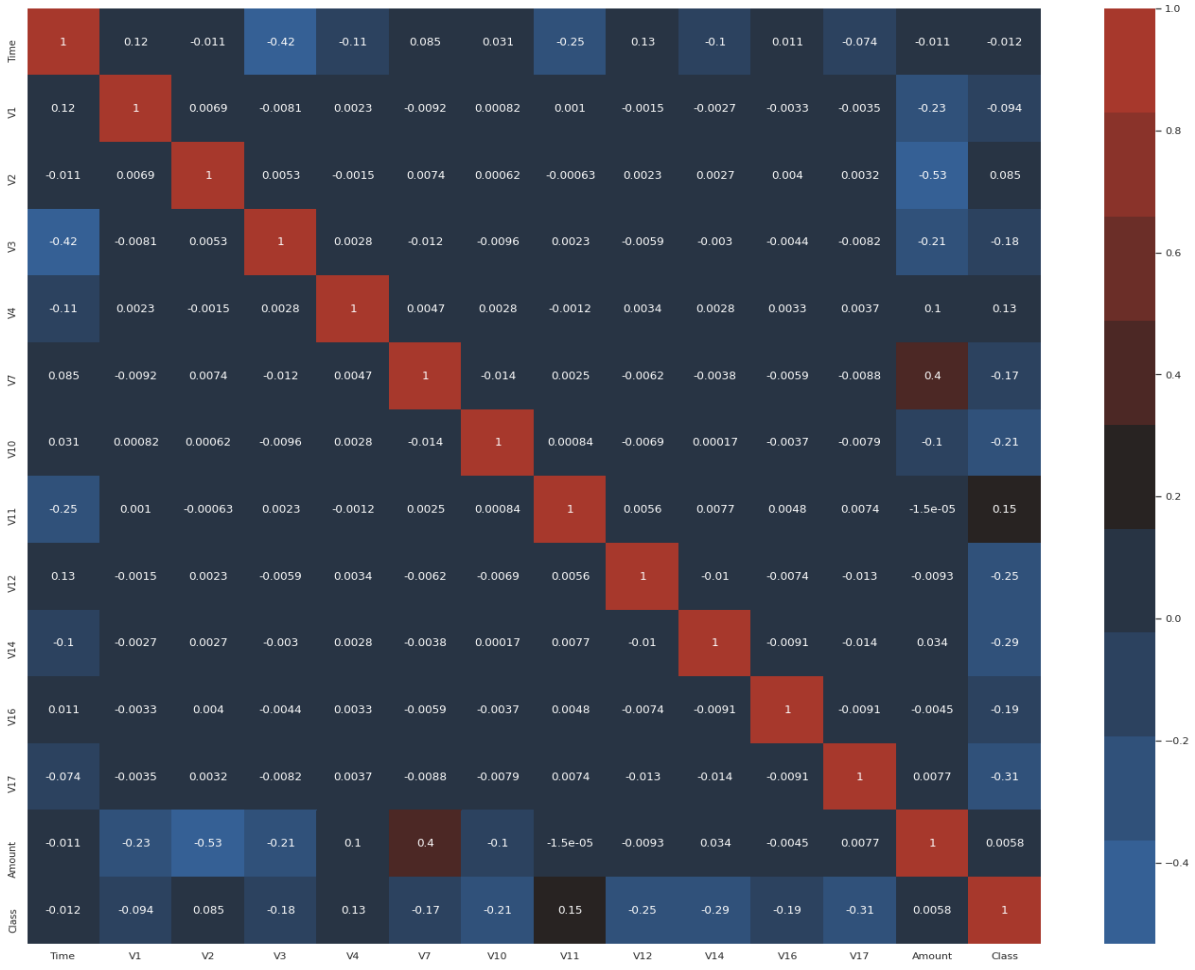


Figure 4: Heatmap representation of a dataset

Table 2: Summary of details of evaluation metrics for the implemented ML algorithms

| Classifiers | Accuracy % | Specificity % | Precision % |
|---|---|---|---|
| Random Forest | 99 | 98.7 | 99.7 |
| Logistic Regression | 100 | 97.9 | 99.6 |
| SVM | 100 | 98.4 | 78.2 |
| Decision Tree | 100 | 91.2 | 91 |
| KNN | 100 | 97.1 | 41.0 |
| Proposed Algorithm | 99.74 | 100 | 100 |

## 5. CONCLUSION

This provides a complete alignment of frameworks for effective fraudulent transaction identification. Initially, it starts with the use of methodologies for clustering and outlier detection. Such methodologies are the basis of data collection that does not match the modern data framework. These are also made precise by the use of ML. Being a binary classification framework, ML serves to give the consumer the outcome of whether the present payment is lawful or unlawful. However, the false-positive occurrences in machine learning methodologies are significant; an alternative standardized animal behavior technique is then implemented to produce further detailed outcomes. The most significant role that a bank may give its clients is to identify the unlawful method. Although if the transaction identified as illegitimate by the machine happens to be legal, that scenario could claim to be a significant drawback. This might lead to a decline in the company's goodwill. Therefore the elimination of false positives is a mandatory feature that any device can implement. The accuracy score hits 100 percent in this method. This system would be capable of distinguishing ordinary purchases from unlawful purchases effectively by being learned from sufficient training data. Yet distorted knowledge from training contributes to less accurate identification. It would then be more effective to build a method that would have improved outcomes by training on regular data to identify illegitimate inputs. The current method can be combined with a classification methodology that evaluates and interprets the accessible data set from illegitimate inputs. If the result is in a human-readable form, the mechanism should be more optimized since the real result from a machine learning-based result is not humanly explicable in existence.

## REFERENCES

[1] Yu, Wen-Fang, and Na Wang. "Research on credit card fraud detection model based on distance sum." 2009 International Joint Conference on Artificial Intelligence. IEEE, 2009.

[2] Ounacer, S., El Bour, H. A., Oubrahim, Y., Ghoumari, M. Y., &Azzouazi, M. "Using Isolation Forest in anomaly detection: The case of credit card transactions" Periodicals of Engineering & Natural Sciences, 6(2), 394–400, 2018.

[3] Benchaji, I., Douzi, S., & El Ouahidi, B. "Using a genetic algorithm to improve classification of imbalanced datasets for credit card fraud detection". Lecture Notes in Networks & Systems, 66, 220–229, 2019.

[4] Xuan, Shiyang, Guanjun Liu, Zhenchuan Li, Lutao Zhang, Shuo Wang, and Changjun Jiang. "Random forest for credit card fraud detection." In 2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC), pp. 1-6, 2018.

[5] Zheng, L., Liu, G., Yan, C., & Jiang, C. "Transaction fraud detection based on total order relation & behavior diversity". IEEE Transactions on Computational Social Systems, 5(3), 796–806, 2018

[6] Maniraj, S. P., Aditya Saini, Swarna Deep Sarkar, and Shadab Ahmed. "Credit Card Fraud Detection using Machine Learning and Data Science". International Journal of Engineering Research & Technology(IJERT), 8(9), pp. 110 – 115, 2019.

[7] Makki, S., Assaghir, Z., Taher, Y., Haque, R., Hacid, M. S., &Zeineddine, H. "An Experimental Study With Imbalanced Classification Approaches for Credit Card Fraud Detection". IEEE Access,7, 93010–93022, 2019.

[8] Maes, Sam, Karl Tuyls, Bram Vanschoenwinkel, and Bernard Manderick. "Credit card fraud detection using Bayesian and neural networks." In: Proceedings of the 1st international naiso congress on neuro-fuzzy technologies, pp. 261-270. 2002.

[9] Fu, Kang, Dawei Cheng, Yi Tu, and Liqing Zhang. "Credit card fraud detection using convolutional neural networks." In: International Conference on Neural Information Processing, pp. 483-490, 2016.

[10] Roy, Abhimanyu, Jingyi Sun, Robert Mahoney, Loreto Alonzi, Stephen Adams, and Peter Beling. "Deep learning detecting fraud in credit card transactions." In: 2018 Systems and Information Engineering Design Symposium (SIEDS), pp. 129-134, 2018.

[11] C. Phua, D. Alahakoon and V. Lee, "Minority report in fraud detection". ACMSIGKDD Explorations Newsletter, 6(1), pp. 50, 2004.

[12] S. Mittal and S. Tyagi, "Performance Evaluation of Machine Learning Algorithms for Credit Card Fraud Detection". 2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2019.

[13] S.Dutt, A.K.Das, and S.Chandramouli, Machine Learning. Pearson Education India, 2018.

[14] Altiti, O. (2020). Credit Card Fraud Detection Based on Machine and Deep Learning. (Section IX), 204–208.

[15] Pande, Sagar, Aditya Khamparia, Deepak Gupta, and Dang NH Thanh. "DDOS Detection Using Machine Learning Technique." In Recent Studies on Computational Intelligence, pp. 59-68, 2021.

[16] Pande, Sagar, and Ajay B. Gadicha. "Prevention Mechanism on DDOS Attacks by using Multilevel Filtering of Distributed Firewalls." International Journal on Recent and Innovation Trends in Computing and Communication 3, (3), pp.1005-1008, 2020.

[17] Pande, Sagar Dhanraj, and Aditya Khamparia. "A Review on Detection of DDOS Attack Using Machine Learning and Deep Learning Techniques", Think India Journal, pp. 2035-2043, 2019.

[18] Khamparia, Aditya, Sagar Pande, Deepak Gupta, Ashish Khanna, and Arun Kumar Sangaiah. "Multi-level framework for anomaly detection in social networking." Library Hi Tech, 2020.

[19] Ganorkar, Shaunak S., Shilpi U. Vishwakarma, and Sagar D. Pande. "An Information Security Scheme for Cloud-based Environment using 3DES Encryption Algorithm." International Journal of Recent Development in Engineering and Technology 2, (4), 2014.

[20] Divya, K., Akash Sirohi, Sagar Pande, and Rahul Malik. "An IoMT Assisted Heart Disease Diagnostic System Using Machine Learning Techniques". Cognitive Internet of Medical Things for Smart Healthcare, pp. 145-161, 2021.

## BIOGRAPHIES

**Ms. Vaishnavi P. Mahalle-** Btech Student
Dept. CSE ,DES's College of Engineering and Technology , Dhamangaon Rly. Maharashtra, India.
Domain : Machine Learning

**Ms. Yashashri G. Rathi-** Btech Student
Dept. CSE ,DES's College of Engineering and Technology , Dhamangaon Rly. Maharashtra, India.
Domain : Machine Learning

**Ms. Shivani P. Pote-** Btech Student
Dept. CSE ,DES's College of Engineering and Technology , Dhamangaon Rly. Maharashtra, India
Domain : Machine Learning

**Ms. Khushboo P. Shriwas-** Btech Student
Dept. CSE, DES's College of Engineering and Technology , Dhamangaon Rly. Maharashtra, India
Domain : Machine Learning

**Kalyani Gholap -** Assistant Professor
Dept. CSE, DES's College of Engineering and Technology, Dhamangaon Rly. Maharashtra, India
Domain: Machine Learning