IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 11, Issue 5, May 2023

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105303 |

Credit Card Fraud Detection Using Machine Learning

Prof. Radha Shirbhate, Somnath Borude, Vishal Hadke, Sushama Walunjkar, Shubham Bhujade,

Department of AI & DS, GHRCEM, Wagholi, Pune, India

Department of IT, GHRIET, Wagholi, Pune, India

Department of IT, GHRIET, Wagholi, Pune, India

Department of IT, GHRIET, Wagholi, Pune, India

Department of IT, GHRIET, Wagholi, Pune, India

ABSTRACT: Today, online business has become an important and necessary part of our lives. Credit card fraud is the most common problem in the world today. This is due to the rise of online business and e-commerce platforms. Credit card fraud often occurs when a credit card is stolen for an illegal purpose or even when a fraudster uses credit card information for their own use. The higher the transaction frequency, the higher the number of scams. This article discusses machine learning algorithms such as logistic regression, decision trees, and random forest classification to reduce fraud. Implement and test the same set of algorithms using online data. From the comparative analysis, it can be concluded that logistic regression, decision tree and random forest classification algorithms perform better in fraud detection.

KEYWORDS: Credit card, Fraud detection, Machine learning, supervised learning, Logistic regression, Decision Tree, Random Forest Classification.

I. INTRODUCTION

Government departments, commercial companies, financial institutions and many other organizations due to the rise and rapid growth of e-commerce. In today's world, credit cards are widely used for online purchases, which has led to many creditcard related fraud cases. In the digital age, the need to verify credit cards is necessary. The main purpose of is fraud detection, which can detect fraud in a shorter time and more accurately, using machine learning-based classification algorithms.

Advances in technology have reduced cash payments and increased online payments, paving the way for scammers to conduct their transactions anonymously.Some online payment methods only ask for card number, expiry date and cvv, this information can be lost without us and sometimes we don't know if our information is stolen. Fraud detection involves monitoring and analysing the behaviour of multiple users to predict the detection or prevention of malicious behaviour. To effectively detect credit card fraud, we need to understand the various techniques, algorithms, and methods involved in credit card identification.

Types of Frauds:

1. Online and Offline
2. Card Theft
3. Data phishing
4. Application Fraud
5. Telecommunication Fraud

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105303 |

II. RELATED WORK

Any form of fraud is a crime and a violation of the law, and credit card fraud is money laundering. There are many studies that try to figure out whether the business is fraudulent. There are still many challenges and attempts to overcome them. Afterconducting a survey of various credit card fraud methods, we can conclude that machine learning itself has many ways to detect credit card fraud.

In 2020,RuttalaSailusha, V. Gnaneswar, R. Ramesh, G. Ramakoteswara Rao discuss the performance of Random Forest and Adaboost algorithms.They took a very skewed dataset and worked on this type of dataset. The logistic regression algorithm is similar to the linear regression algorithm. Linear regression is used to estimate or predict values. The Adaboost algorithm classifies transactions as fraudulent and non-fraud transactions as fraudulent. The algorithms used are random forest algorithm and Adaboost algorithm.The results of both algorithms are based on accuracy, precision, recall and F1score. The ROC curve is drawn from the confusion matrix. When Random Forest and Adaboost algorithms are compared, the algorithm with the highest accuracy, precision, recall and F1 score is considered the best algorithm for fraud detection. From their study, they concluded that both Adaboost and Logistic regression achieved the highest accuracy.

In 2021, D. Tanouz, .R Raja Subramanian, D. Eswar, GV Parameswara Reddy, A. Ranjith Kumar, CH V N M praneeth explain the function of decision trees, random forest logistic regression, naive Bayes classification.It is one of the Naive Bayes classification algorithms. The algorithm is based on Bayes' theorem. Bayes' theorem states that the probability of an event is given. The logistic regression algorithm is similar to the linear regression algorithm. Linear regression is used to estimate or predict values.Logistic regression is often used for classification functions. The J48 algorithm is used to construct decision trees in classification problems. J48 is the continuation of ID3 (Iterative Dichotomy). J48 is one of the most used and analysed areas in machine learning. The algorithm mainly works with continuous and categorical variables.They concluded that the random forest algorithm has the highest accuracy among other algorithms and is considered the best algorithm for fraud detection. The Random Forest classifier performed best, with 96.7741% accuracy, 100% precision, 91.1111% recalls, and 95.3488% f1 scores for 95.5555 ROU-AUC points.

III. METHODOLOGY

A . Dataset

Dataset is from the ULB Machine Learning Group. Data includes European cardholders' credit card purchases in September 2013. The data shows transactions that took place over two days and includes 284,807 transactions. The positive class (fraud) accounts for 0.172% of data transfer.The data is very unstable and biased towards quality classes. It contains only numerical (continuous) input variables that are the result of 28 principal components generated by the principal component analysis (PCA) feature selection transformation. Therefore, a total of 30 strategies were used in this study. Due to privacy concerns, details and background information about the features cannot be provided. The physical property includes the number of seconds since each transaction and the first change in the database.

The "Amount" attribute is the exchange rate. The unique "class" is intended to be a binary class with a value of 1 for positive events (fraud) and 0 for negative events (non-fraud).

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105303 |

B. Hybrid Sampling of dataset

Create a preliminary file of the dataset. A combination of under-sampling and over-sampling is performed on quite different data to obtain two sets of distributions for analysis. This is done by incrementally adding and subtracting interpolated data points from existing data points until overfitting is reached. During under-sampling, we generate normal data equal to the fraud data. During oversampling, we create dummy data similar to fraud data.

C. Logistic Regression Classifier

Logistic regression using a functional method to estimate the probability of a binary response based on one or more variables (traits). It found a best fit to a function called the sigmoid. Sigmoid function (σ) and introduction to sigmoid function (x). The purpose of using logistic regression is to find the most appropriate model that explains the relationship between dependent and independent variables. Despite name regression, LR is used in the distribution of problems estimating binomial and polynomial results where the purpose is to estimate the value of parameter coefficients using the sigmoid function. Logistic regression is used for integration, when a transaction occurs it checks its value and decides whether to continue the transaction or not. Results from the completed procedure give approximately 94.48% accuracy.

D. Decision Tree Classifier

Decision trees are supervised learning techniques that can be used for both classification and regression problems, but are mostly used to solve classification problems. It is a representation to get all the solutions of the problem/decision based on the given situation. In a decision tree, the algorithm starts at the root of the tree to predict the class of a given data. The algorithm compares the value of the root attribute with the value of the data (actual dataset) attribute and follows the branch to jump to the next node based on the result of the comparison. It is a calculation tool for classification and prediction. A tree has internal nodes that represent a test of aattribute, each branch represents that outcomes of test, and each leaf node (terminal node) contains a class label. Results from data processing give an accuracy of about 99.83%.

E. Random Forest Classifier

The random forest algorithm belongs to the category of learning algorithms. We create N decision tree models in the random forest algorithm. Each model predicts the target value. Using the voting method usually estimates the final value of the target. In the proposed method, we use the Random Forest Algorithm (RFA) to find the fraudulent transactions and the accuracy of those transactions. The algorithm is based on a supervised learning algorithm in which a decision tree is used to classify the data. After dividing the dataset you get the confusion matrix. Evaluate the performance of the random forest algorithm based on the confusion matrix. Results from data processing provide up to 99.99% accuracy. We deploy the model on random forest classifier because we get the higher accuracy for random forest classifier.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105303 |

ARCHITECTURE:

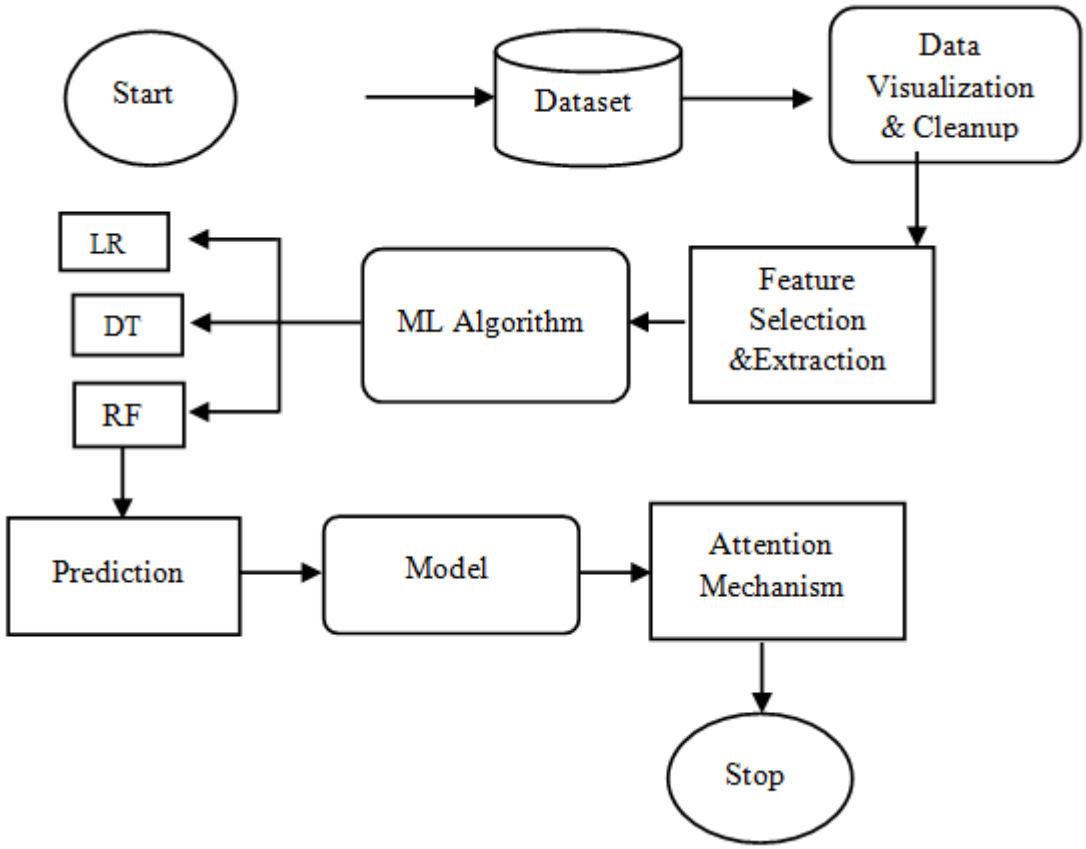

Fig . Architecture of credit card fraud detection system

System Implementations Plan:

The fraud detection module will work in the following steps:

1. Upload the dataset which has the transactions and amount are treated as credit card transactions.
2. After dataset upload the feature selection and dimensionality reduction step is done.
3. We separate out the dataset as a train dataset (80%) and test dataset (20%) .
4. The train dataset are given to machine learning algorithms as an input.
5. We train the model using this training dataset.
6. After that we test the model using testing dataset.
7. The valid transactions are treated as genuine transactions.
8. And fraudulent transactions are treated as Fraud transaction.
9. After that we check the accuracy.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105303 |

PROJECT MODULE:

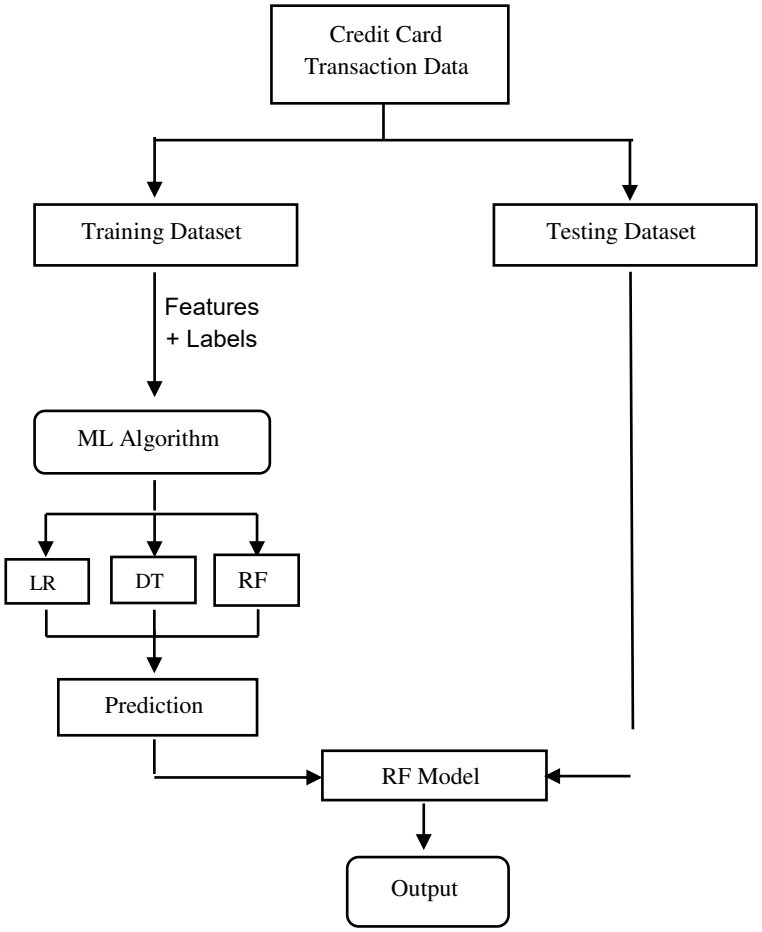

Fig. Project Module

In project module we can see that, we use the three different algorithms logistic regression, decision tree and random forest classifier. We check the accuracy of each algorithms and after that compare the result of these all. We get the higher accuracy for random forest classifier which is 99.99%.

III. RESULTS

a) Using Under sampling:

Sr.No.	Models	Accuracy
1	LR	93.15
2	DT	88.94
3	RF	93.15

b) Using Over Sampling:

Sr.No.	Models	Accuracy
1	LR	94.48
2	DT	99.83
3	RF	99.99

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105303 |

Here we can see that we get the higher accuracy from over sampling technique. We can see that the Random Forest Classifier has highest accuracy from all of the models , that's why we save the model of random forest classifier.

Advantages:

1. Faster detection
2. Higher Accuracy
3. Improved efficient with larger data
4. Fewer false declines
5. Less manual work needed for additional verification
6. Ability to identify new patterns and adapt to changes.

IV. CONCLUSION

An effective credit card fraud detection system is a top requirement for any credit card access merchant, but it can cash out millions of credit cards online by fraud. Credit card fraud is undoubtedly an unethical crime. Credit card analysis hopes to help people avoid bankruptcy. Machine learning is preferred for fraud detection due to its high accuracy and cost. Researchers are still working to achieve higher accuracy and detection rates. We use random forest, logistic regression and decision tree classifiers for fraud detection. We save the random forest model because it gives you better people than logistic regression and decision tree classifiers. Provides greater accuracy of higher data.The Random forest algorithm will perform better with a larger number of training data, but speed during testing and application will suffer. Application of more pre-processing techniques would also help.

REFERENCES

[1] Proceedings of the International Conference on Intelligent Computing and Control Systems (ICICCS 2020) IEEE Xplore Part Number: CFP20K74-ART; ISBN: 978-1-7281-4876-2
[2] Proceedings of the Fifth International Conference on Intelligent Computing and Control Systems (ICICCS 2021) IEEE Xplore Part Number: CFP21K74-ART; ISBN: 978-07381-1327-2
[3] Proceedings of the 2nd International Conference on Trends in Electronics and Informatics (ICOEI 2018) IEEE Conference Record: # 42666; IEEE Xplore ISBN:978-1-53863570-4
[4] Chaudhary, K. and Mallick, B., (2012). Credit Card Fraud: The study of its impact and detection techniques, International Journal of Computer Science and Network (IJCSN), Volume 1, Issue 4, pp. 31 – 35, ISSN: 2277-5420
[5] RamaKalyani, K. and UmaDevi, D., (2012). Fraud Detection of Credit Card Payment System by Genetic Algorithm, International Journal of Scientific & Engineering
Research, Vol. 3, Issue 7, pp. 1 – 6, ISSN 2229-5518
[6] N. Mahmoudi, E. Duman, "Detecting credit card fraud by Modified Fisher Discriminant Analysis", Elsevier Expert System with Application, 2015, pp. 2510-2516.
[7] N. Halvaiee, M. Akbari, "A novel model for credit card fraud detection using Artificial Immune System", Elsevier Applied Soft Computing, 2014, pp. 40-49.

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.379

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details