WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Credit Card Fraud Detection using Machine Learning and Data Science

Akshada dighe[1], Shantanu Patil[2], Pratik Thakre[3], Gauri Tupe[4], Pranav Kenjale[5]

([#1] Assistant Professor, Department of Information Technology, Genba Sopanrao Moze College of Engineering)
[(#2, 3, 4, 5] BE IT Students , Genba Sopanrao Moze College of Engineering)

ABSTRACT-

The use of credit cards is over now these days. It is important for credit card companies to be able to detect credit card fraud so that customers do not buy with items they did not purchase. These problems can be overcome by using Machine Learning. The purpose of this activity is to detect credit card fraud in less time and with greater accuracy. This model is used to identify whether the activity is counterfeit or real. Machine learning algorithms are used to analyze what is being done and report suspicions in a timely manner. Our goal is to detect 100% fraud in a short period of time with greater accuracy.The credit card fraud program falls under Supervised Learning. In this program, we should focus on analyzing and analyzing data sets. For this, we use algorithms such as Random Forest Classifier, AdaBoost Classifier, CatBoost Classifier, and XGBoost Classifier. Scikit-learn is mostly written python, and we use Numpy for fast line algebra and array functions, and we use different algorithms used in the sklearn library. Everything the research is done in the language of the python program.

Keywords-- Credit Card Fraud, Machine Learning, Scikit-Learn, Random Forest Algorithm, AdaBoost Algorithm, CatBoost Algorithm, XGBoost Algorithm.

Introduction-

In today's world, economic losses are increasing rapidly. There are many hijackers around the world because of innocent people being deceived. Hackers are constantly trying to find new laws and strategies to carry out illegal activities.

A team of Data Scientists, Data Analysts are constantly working on it and finding ways to detect fraud and keep users safe from these losses. They use the techniques of Data mining, data analysis, machine learning to find effective solutions to these problems. They use a lot of machine learning, math, and artificial intelligence techniques.

In our project to detect credit card fraud, we have used a number of machine learning algorithms such as Bagging and Boosting. Our main goal is to compare 4 categories and find the fastest and most accurate. So, basically, we have a web application with an ML model running in the background. Through our app, we have trained the best model with 2 columns namely; variable it is a class that contains 0 or 1. Class 0 represents the actual function and 1 represents the counterfeit function.

The second most important part of our project is the machine learning model. We have demonstrated the use of 4 models where 1 bagging and the other 3 boosting algorithms. The reason for using the 1 bagging algorithm is because the paper we are referring to has given us the conclusion that the Random Forest Classifier if the integrated learning method is the most accurate. Yes, there is no doubt that it is more accurate and its accuracy is the best of all, but the second major factor is the totally unacceptable timing. Accuracy should be high but the time required should be minimal. When we compared the performance time between Random Forest and 3 boosting algorithms, we found that it was the slowest. Our job is to produce a fast and accurate app. Therefore, we have selected the Catboost Classifier as the best because it is accurate and the time required is very less compared to **Random Forest Classifier.**

The main purpose of the project is to advance the research project using new algorithms one step further. This project clearly shows which algorithm is best and should be used for real-time fraud detection. This Credit Card project uses Catboost Classifier and provides accurate and fast results.

Literature review:

Many books on mysterious discovery or fraud in this domain have been published and are available for public use. Learning to detect credit card fraud. Although these methods and algorithms have achieved unexpected success in some places, they have failed to provide permanent and consistent fraud detection solutions.

This process has been proven to be effective in reducing false alarm levels as well increase the rate of fraud detection. The author reads different ways to get foreign objects. Use these how to detect fraud. Credit card fraud can be solved using these external methods. These methods are useful for finding fraudulent jobs.

Three differentiating models are supported in the Neural network, the closest k-logistic neighbor retreat is enhanced. To test these models, 70% of the

database is used for training while 30%. set aside for verification and testing. Accuracy, sensitivity, clarity, accuracy, Matthew's correlation coefficient, and balance Divide rate is used for estimation three-phase operation

Algorithm for obtaining data stream based on an anti k neighbor nearby to get a loan card fraud detection while traditional methods need to scan the database multiple times to find it fraudulent sales, misappropriated data surrounding streams. This method is easy to stop fraudulent work is lost and credit is stolen card. Confirmation tests and finding errors in a sequences number also help to find easy and invalid numbers easily.

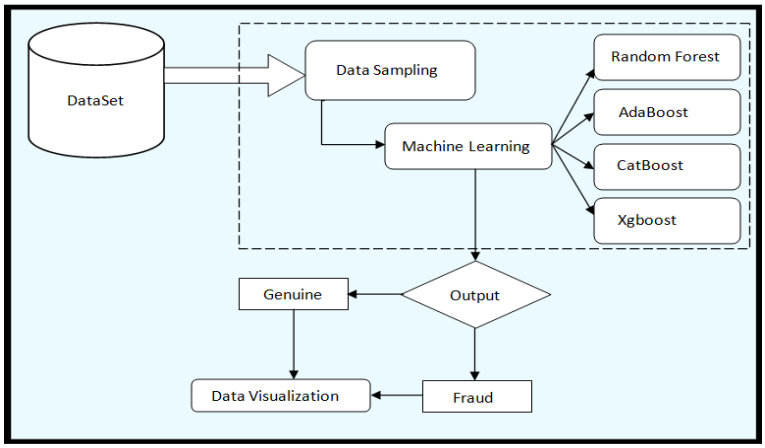
The increase in online transactions is directly related to an increase in the number of frauds. In this paper variety of Algorithms like K-Nearest Neighbor, Rainforest, AdaBoost, cat boost, Xgboost, and Logistic Regression are some challenges involving splitting normal transactions and frauds appear to be very similar to each other. Limitations for obtaining these functions Time, Number, and Frequency of Activity. In this paper, the four K-Nearest Neighbor are different algorithms AdaBoost, Rainforest, and Logistic Decrease compared to fraud detection methods. The logistic decline is relatively good in other algorithms. This model uses unequal credit card fraud data. All of these algorithms do not work immediately to detect fraud in operations.

Proposed methodology:

It is always difficult to work to detect fraud transactions using previous transactions data with the help of traditional methods, and 99% of users don't report fraud so which makes it more difficult to trace it. Traditional methods we are using for a long time to detect fraud transactions are time-consuming and not sufficient to detect fraud transactions. We need a system that is capable to detect fraud transactions by analyzing previous data.

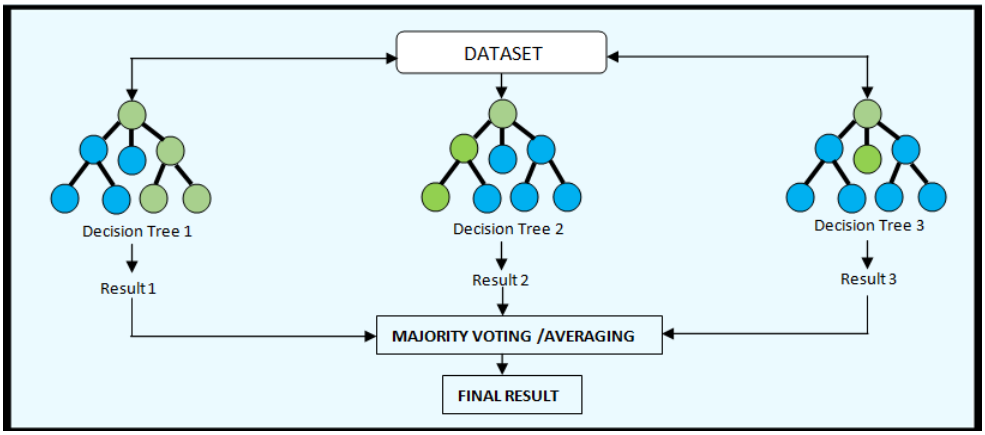
The method proposed by this paper uses the latest machine learning algorithms to detect missing amounts and external factors.

Our model uses early data processing techniques using feature selection and minimizing the size of the credit card fraud set, to reduce the number of input features prior to inclusion in the model. The student's short-term memory consecutive memory is then used as a flexible pattern recognition component to capture consecutive interdependence between consecutive credit card purchases. Next, an attention-grabbing approach was introduced to provide a unique focus on information that emerges from the hidden layers of short-term memory, allowing our model to find the right fraud patterns and find the best jobs that are very different between consumer purchases. In the proposed system, we are using different machine learning algorithms to analyze the data and to get desired output.

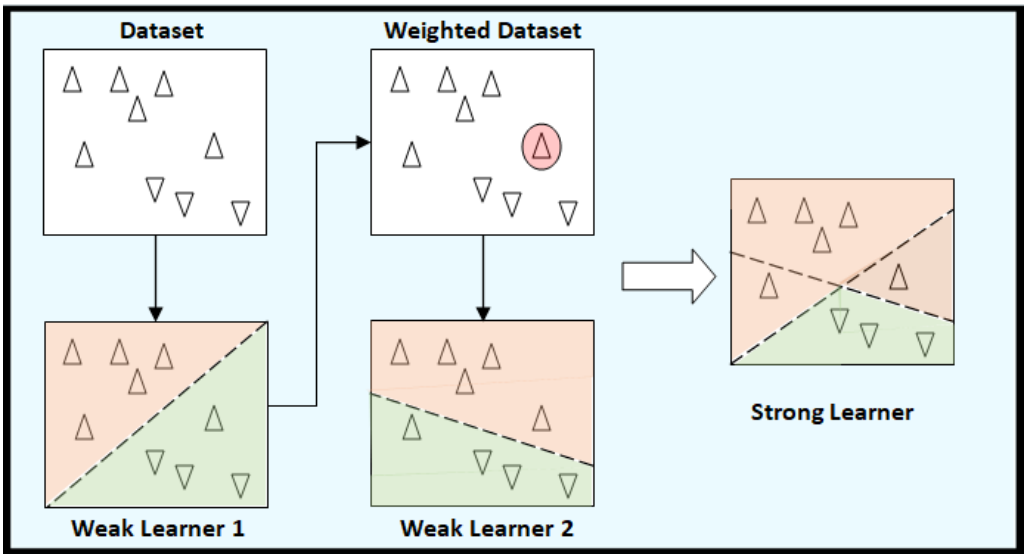


This system consists of four algorithms: one for bagging and the other three for boosting.

1) Random forest algorithm: Random Forest is a supervised machine learning algorithm that is widely used for classification and regression problems. Build decision-making trees from different samples and take their majority vote to take classification and average in the event of a recession. One of the most important features of the Random Forest Algorithm is that it can manage a set of data containing continuous variables as in the case of regression and categorical variables. It produces the best results for classification problems.



2) Adaboost: AdaBoost is the most representative algorithm in the Boosting family. It saves I distribution of one set of training sample opportunities and fixes this Distribution of opportunities for each sample during each multiplication. Direct reading the algorithm is used to generate a member separator and calculate its error rate training samples. AdaBoost will use this error rate to correct the chances distribution of training samples. The role of weight change is a major set-up wrongly weighted sample and reduce its weight if the sample is so properly classified. Finally, by way of a weighted vote of individual separatists, a strong phase will be established.



3) Catboost: CatBoost is an algorithm for increasing the gradient in decision trees. Developed by Yandex researchers and engineers, it follows the MatrixNet algorithm widely used within the company to measure tasks, predict and make recommendations. It is available worldwide and can be used in many different areas and problems. Catboost gets the best results from benchmarking, and that's great.

4) Xgboost:The XGBoost classifier is a tree-based machine learning algorithm, which uses a gradient boost framework. It was used to solve regression, classification, ranking, and prediction problems defined by the user.It is a complete combination of software and hardware techniques to achieve high results using minimal computer resources in a very short time.

Boosting Methods	Accuracy	Time
AdaboostClassifier	98.3033	156
XGBoostClassifier	98.9526	123
RandomForestClassifier	99.9894	460
CatBoostClassifier	99.9529	152

Conclusion –

In this survey we successfully Studied various algorithms XGBoost classifier, CatBoost classifier, AdaBoost classifier, Random Forest Classifier, these four algorithms, using sci-kit learn python, This survey gives a clear idea about the algorithms, which is much better, faster, and accurate for model training, in the field of machine learning and data science. The purpose of this study is to explore as many methods that can detect fraud effectively. Replacing the old algorithms with new effective ones is always a better option.

References

- "Credit Card Fraud Detection using Machine Learning Algorithms" by Vaishnavi Nath Dornadulaa , Geetha S published by INTERNATIONAL CONFERENCE ON RECENT TRENDS IN ADVANCED COMPUTING 2019, ICRTAC 2019
- "Credit Card Fraud Detection Using Machine Learning And Data Science" by Bhargava R1 , Ajay Kumar K2 , Bhavana R3 , Sai Charan S4 , S Lokeswara5
- "Machine Learning Model for Credit Card Fraud Detection- A Comparative Analysis" by Pratyush Sharma, Souradeep Banerjee, Devyanshi Tiwari, and Jagdish Chandra Patni School of Computer Science, University of Petroleum and Energy Studies Dehradun published by The International Arab Journal of Information Technology, Vol. 18, No. 6, November 2021

- "Credit Card Fraud Detection using Machine Learning and Data Science" by Aditya Saini, Swarna Deep Sarkar Shadab Ahmed Department of Computer Science and Engineering SRM Institute of Science and Technology published by International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181 Vol. 8 Issue 09, September-2019
- "Review on Credit Card Fraud Detection using Machine Learning Algorithms" by Pooja, Dr. Ashlesha-J.C. Bose University of Science and Technology YMCA at Faridabad, Haryana - India published by International Journal of Computer Trends and Technology (IJCTT) – Volume 68 Issue 6 – June 2020