iJRASET
International Journal For Research in Applied Science and Engineering Technology

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 11 **Issue:** II **Month of publication:** February 2023

**DOI:** https://doi.org/10.22214/ijraset.2023.49213

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538*
*Volume 11 Issue II Feb 2023- Available at www.ijraset.com*

# A Study of Machine Learning Effectiveness in Detecting Credit Card Fraud: KNIME

Prathibha T [1], Dr. Arjun B. C[2]
[1] *M.Tech Student, Rajeev Institute of Technology, Hassan*
[2] *HOD, ISE, Rajeev Institute of Technology, Hassan*

***Abstract:*** ***Credit cards are becoming the most widely utilized form of payment. The numbers of fraud users are growing as quickly as the technology. This paper discusses the performance of three popular Machine Leaning techniques for predicting credit card fraud detection. In this paper we used Machine Learning algorithms such as Decision Tree, Random Forest and Simple Regression tree for handling the imbalanced credit dataset. KNIME is used as data analytical tool for the purpose of simulation. The Comparative Study related to Accuracy is being tested and the Regression tree will give the best result.***
***Keywords: Machine Learning, KNIME, Classifier, Decision tree, Random Forest, .Regression Tree***

## I. INTRODUCTION

Identifying fraud credit card transactions is the main goal of this paper. In order to achieve this we need to do classification of the fraud and valid transactions from the dataset. Fraud detection is an example of an anomaly detection task where we can identify the third party activity.

The objective is to create a fraud detection model that uses machine learning-based classification methods to identify an unusual transaction with high accuracy and low error rate.

By modelling the purchasing habits of user, a credit card company can detect misuse of owner cards. If a thief steals credit card or credit card information, the thief's purchases will be the different probability than the original one,. By analysing the different probability, we can identify the invalid transaction. To do this many machine learning and deep learning techniques are there to detect the fraud transactions. In this paper, we are using machine learning classification algorithm such as Random forest, Decision tree and simple regression tree algorithm for analysing the imbalanced credit card dataset and we are analysing its performance through the accuracy with low error rate, which gives high accuracy

## II. LITERATURE SURVEY

In earlier studies, data mining techniques were used to identify fraud using a conventional approach. Nuno Carneiro, Goncalo Figueira, Miguel Costa,[1]has discussed the limitations in the A data mining based system for credit card fraud detection in e-tail", Elsevier,pp.91- 101. In this paper we are using Machine learning based supervised classification algorithms on the dataset and the dataset has been taken from kaggle[2] website..Akinyelu and O. Adewumi[3] conducted a detailed study on fraud detection using the method of natural observation of customer-side events. The Nilsson Report [4] provides a detailed report on the various methods of fraud or scam presence in the field of credit card business, as well as the various methods of identifying them and the adverse effects of the scams on business environments'.

For implementing the Decision tree with proper classification Y. Sahin, S. Bulkan, and E. Duman[5] , " had conducted a detailed study on A cost -sensit ive decision tree approach for fraud detection," Expert Systems with Applications, vol. 40, no. 15, pp. 5916–5923, 2013 which gives an idea about classification

Research scientist Ian Good fellow [6] works for OpenAI there he discussed the importance of learning algorithms, supervised algorithms along with the various probability distributions where time is given as main criteria for the validation purpose

## III.RESEARCH FINDINGS

The primary objective of this study is to detect the fraud invalid transactions with low error rate using machine-learning algorithms.

## IV.PROPOSED METHODOLOGY

In this project three machine-learning algorithms namely Decision tree, Random Forest and Simple Regression Tree classification algorithms techniques are applied on real European data set

For implementation of any algorithm or techniques in general there are some certain steps need to be considered.

*1)* Data Collection Phase: It is the first and very important phase from where the exact process begins. We need to collect the related data, which suits the requirement.
*2)* Data Validation Phase: In this phase, we will check whether the data we have collected will exactly related to our application.
*3)* Data Analysis Phase: In this phase, we analyse the collected data by implementing it in various models in various criteria's.
*4)* Data Reporting: It involves the reporting of data, which have been collected from the previous stage.

For Data collection we are using publicly available dataset downloaded from kaggle named as creditcard.csv.

### *A. About Dataset*

The dataset includes the credit card transactions made by European cardholders in September 2013. The dataset is highly imbalanced one. The dataset contains only the numerical values as the result of a PCA transformation. The features namely Time and Amount is not transformed with the PCA and the other features from v1,v2…v28 are the principal component obtained with PCA

### *B. Implementation Of Algorithms*

#### *1) Using Simple Regression tree*

The algorithm takes the training data from the input port the data will be read by related .csv file. The data will be partitioned into 70% training data and 30% testing data by selecting partition node. In the regression learner node related to configuration node we specify the target column as 'class' and it will be given to predictor node where based on the target value column specified class we will get the predicted output. The output can be seen by using either numeric scorer which gives the predicted output in terms of accuracy and error rate and the simulation is shown in figure 1.

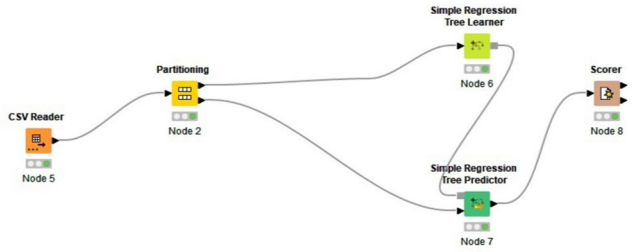



Figure 1. Simple Regression tree workflow

#### *2) Using Random Forest*

To construct random forest we have to train N decision trees. The data will be read from .csv . As the dataset will contain a string value by name class. It has to be converted into numerical data, which will be done by using Number to String Converter, The learner will take the input from the partition and here the output will be seen by Rule engine where we fix the threshold value range as 0.3. By using numeric scorer, we can compute the accuracy in terms of fraud and valid transactions.

Give the related prediction value column for target column Class=0 implies that no fraud and for 1 implies fraud detection. In the rule engine will write the threshold of acceptance for the legitimate class is increased. For this case study, we adopted a decision threshold of 0.3 on the probability of the fraudulent class and compared the results with what we obtained with the default threshold of 0.5. The obtained data will be given to Evaluation, which gives the number of invalid transactions confusion matrix true false. Scorer will give the accuracy statistics. The workflow is shown in figure 2 and the related evaluation is shown in figure 3

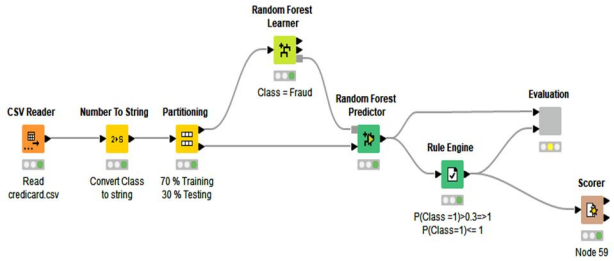



Figure 2. Random Forest Workflow

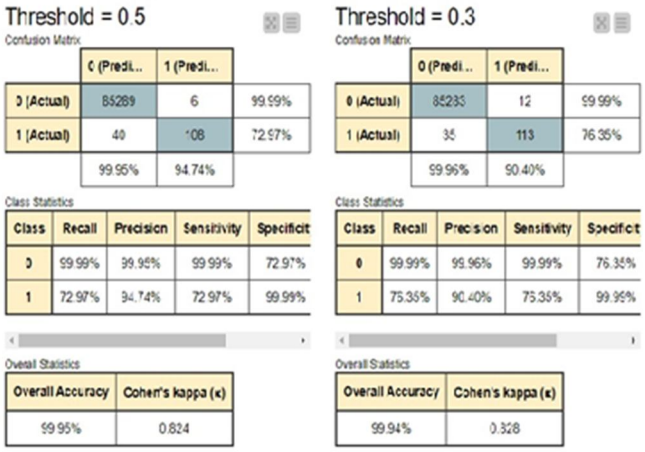



Figure 3 : Evaluation of Random Forest workflow

*3) Using Decision Tree Classifier*

In decision tree we build a tree based on classification of target variable class which is related to the value of 0 and 1. Based on the values the tree will be built. And the dataset partition will be done 70% training and 30% testing and we analyse the performance through the scorer for accuracy.The workflow is shown in figure 4 and the simple decision tree for two catagorical data is shown in figure 5

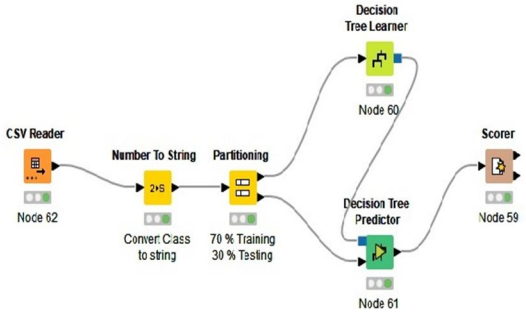



Figure: 4 Decision Tree Classifier

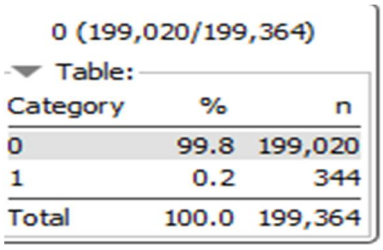



Figure 5: Simple Decision tree view

## V. RESULT ANALYSIS

Three different machine-learning algorithms have been used to detect fraud in credit card transactions, and it has been found that the Regression Tree will provide the highest accuracy and lowest error rate and the graphical representation is shown in Figure 6

Table 1 gives the Comparison of Accuracy Statistics of Linear Regression, Random Forest and Decision Tree tested for different partition

| Algorithms | Accuracy | | |
|---|---|---|---|
| Partition | 80/20 | 70/30 | 60/40 |
| Simple Regression Tree | 99.5 | 100 | 98.9 |
| Random Forest | 99.9 | 99.4 | 98.5 |
| Decision Tree | 98 | 99 | 96 |

Table 1: Comparison of accuracy

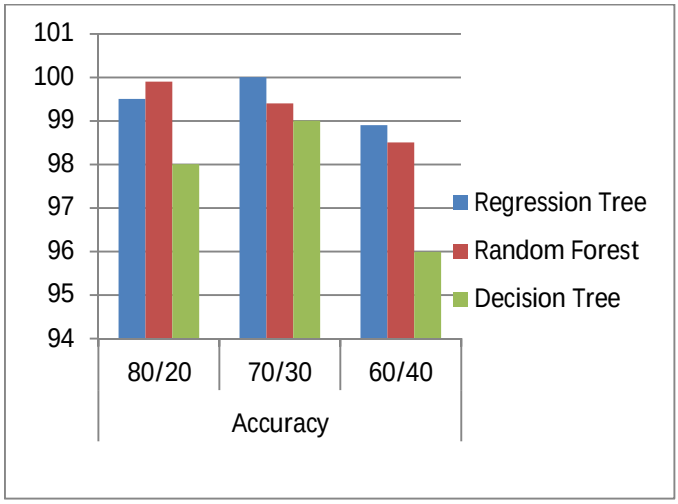


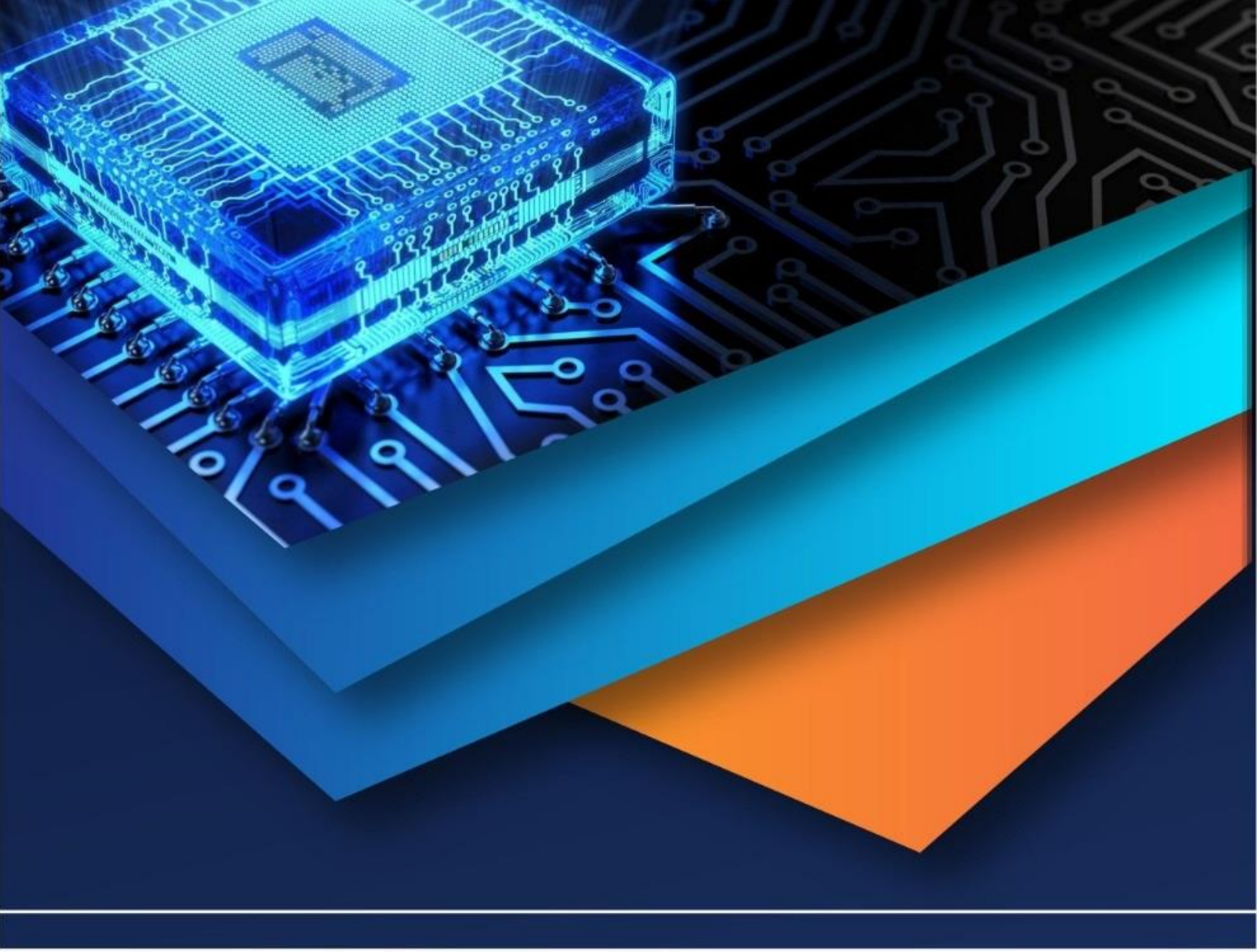

Figure 6. Accuracy classification based on training and testing data

## VI. CONCLUSION

In this paper, we used machine-learning algorithms for detecting the credit card fraud. The model has been properly trained for predicting the occurrence of fraud in the transaction with varying training and testing data units. According to experiment, studies conducted for varying training and testing units Regression Tree will produce the best results in detecting fraud credit card transactions.

## REFERENCES

[1] Nuno Carneiro, Goncalo Figueira, Miguel Costa, “A data mining based system for credit card fraud detection in e-tail”, Elsevier,pp.91- 101
[2] www.kaggle.com/datasets/creditcard.csv
[3] A. Srivastava, A. Kundu, S. Sural, A. Majumdar, “Credit card fraud detect ion using hidden Markov model,” IEEE Transact ions on Dependable and Secure Computing, vol. 5, no. 1, pp. 37–48, 2008K. Elissa, “Tit le of paper if known,” unpublished.
[4] TheNilsonReport(October2016 )[Online].Available:https://www.nilsonreport .com/upload/content_promo/The_Nilson _Report_10-17-2016.pdf [4] Bello-Orgaz, G., Jung, J. J., & Camacho, D. (2016). Social big data: Recent achievements and new challenges. Information Fusion. 28 (Mar. 2016), 45—59
[5] Y. Sahin, S. Bulkan, and E. Duman, “A cost -sensit ive decision tree approach for fraud detection,” Expert Systems with Applications, vol. 40, no. 15, pp. 5916–5923, 2013.
[6] Deep Learning by Lan Good fellow and YoshuaBengiort



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY