Swindling Shonky Anatomization of Credit Card Transactions using Machine Learning

M. Shyamala Devi, Nariboyena Vijaya Sai Ram, Aravapalli Sai Vamshi, Basyam Bharath, Mallangi Surya Prakash Reddy

Abstract: ***With the fast moving technological advancement, the internet usage has been increased rapidly in all the fields. The money transactions for all the applications like online shopping, banking transactions, bill settlement in any industries, online ticket booking for travel and hotels, Fees payment for educational organization, Payment for treatment to hospitals, Payment for super market and variety of applications are using online credit card transactions. This leads to the fraud usage of other accounts and transaction that result in the loss of service and profit to the institution. With this background, this paper focuses on predicting the fraudulent credit card transaction. The Credit Card Transaction dataset from KAGGLE machine learning Repository is used for prediction analysis. The analysis of fraudulent credit card transaction is achieved in four ways. Firstly, the relationship between the variables of the dataset is identified and represented by the graphical notations. Secondly, the feature importance of the dataset is identified using Random Forest, Ada boost, Logistic Regression, Decision Tree, Extra Tree, Gradient Boosting and Naive Bayes classifiers. Thirdly, the extracted feature importance if the credit card transaction dataset is fitted to Random Forest classifier, Ada boost classifier, Logistic Regression classifier, Decision Tree classifier, Extra Tree classifier, Gradient Boosting classifier and Naive Bayes classifier. Fourth, the Performance Analysis is done by analyzing the performance metrics like Accuracy, FScore, AUC Score, Precision and Recall. The implementation is done by python in Anaconda Spyder Navigator Integrated Development Environment. Experimental Results shows that the Decision Tree classifier have achieved the effective prediction with the precision of 1.0, recall of 1.0, FScore of 1.0 , AUC Score of 89.09 and Accuracy of 99.92%.***

Index Terms: ***Machine Learning, Recall, FScore, Accuracy and AUC Score.***

Manuscript published on November 30, 2019.
* Correspondence Author
M. Shyamala Devi, Associate Professor, Computer Science and Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Avadi, Chennai, TamilNadu, India.
Nariboyena Vijaya Sai Ram, III Year B.Tech Student, Computer Science and Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Avadi, Chennai, TamilNadu, India.
Aravapalli Sai Vamshi, III Year B.Tech Student, Computer Science and Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Avadi, Chennai, TamilNadu, India.
Basyam Bharath, III Year B.Tech Student, Computer Science and Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Avadi, Chennai, TamilNadu, India.
Mallangi Surya Prakash Reddy, III Year B.Tech Student, Computer Science and Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Avadi, Chennai, TamilNadu, India.

I. INTRODUCTION

In machine learning, the prediction of fraud credit card transaction is done either by regression or classification process. The entire nation is moving towards online transactions through credit card payment and it is possible due to the technological growth. Fraud online transaction is one of the criminal activities and it must be found at the source itself. This lead to the usage of machine learning approach for the prediction and analysis of transactions.

The paper is organized in which the literature survey is dealt with Section 2 followed by the proposed work in the Section 3. Implementation and the performance analyis is discussed in Section 4 followed by the conclusion of the paper in Section 5.

II. RELATED WORK

A. Literature Survey

The dimensionality reduction can be done by the feature extraction and selection and is considered in predicting the target variable [1]. The general policy regulations, rules and standards are also considered in predicting the target variable [2]. The prediction of the target variable for credit card transaction is done with the classification methods and it is used to categorize the class of transaction [3]. The markov model is used for predicting the fraud credit card and debit card online transaction [4]. The analysis of the whole online credit card data is needed for predicting the online fraud detection and the machine learning approaches can be used to implement this [5]. Several data mining tools and approaches can be used for predicting the credit card fraud detection. The manual computation of detecting the fraud credit card online transaction detection is a tedious and time consuming process and it lead to impractical condition [6]. The fraud in the credit card transaction can be due to inner and outer environment and the fraud may be due to the credit card stole and unusual way of handling the online transaction [7].

The machine learning feature selection and feature extraction methods can be used for the prediction of any factor in different application can be learnt through this article [8] – [21].

III. PROPOSED WORK

In this paper, we have used machine learning classification algorithm for predicting the fraudulent credit card transaction. Our contribution of predicting fraudulent credit card transaction is done in four ways.

(i) Firstly, the relationship between the variables of the dataset is identified and represented by the graphical notations.
(ii) Secondly, the feature importance of the dataset is identified using Random Forest, Ada boost, Logistic Regression, Decision Tree, Extra Tree, Gradient Boosting and Naive Bayes classifiers.
(iii) Thirdly, the extracted feature importance if the credit card transaction dataset is fitted to Random Forest classifier, Ada boost classifier, Logistic Regression classifier, Decision Tree classifier, Extra Tree classifier, Gradient Boosting classifier and Naive Bayes classifier.
(iv) Performance analysis is done by analyzing the performance metrics like AUC Score, Accuracy, FScore, Precision and Recall.

A. System Architecture

The overall design of this paper is shown in Fig. 1

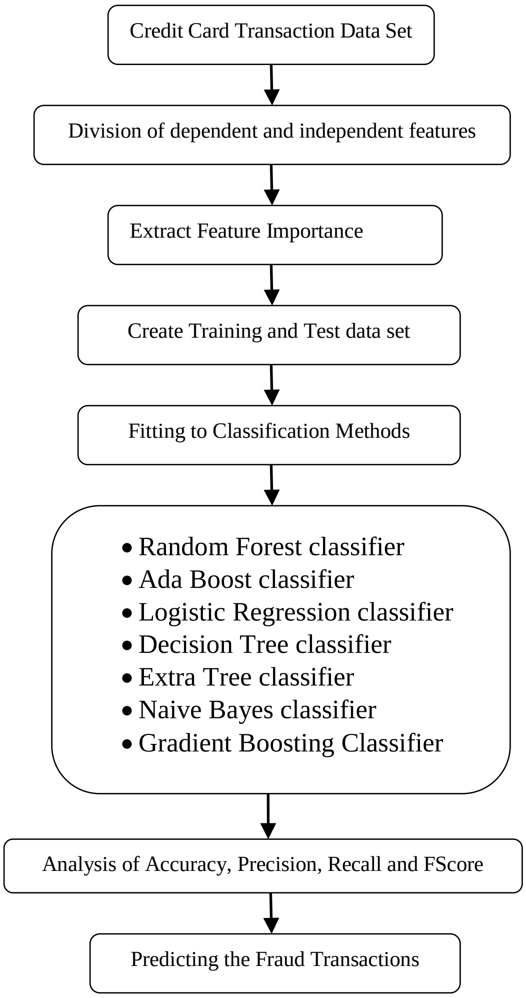

Fig. 1 System Architecture

IV. IMPLEMENTATION AND PERFORMANCE ANALYSIS

A. Data Set Information

The Credit Card Transaction Data Set from KAGGLE Machine Learning Repository is used for execution with 30 independent attribute and 1 Fraud Class dependent attribute and they are as follows,

(1) Time
(2) V1
(3) V2
(4) V3
(5) V4
(6) V5
(7) V6
(8) V7
(9) V8
(10) V9
(11) V10
(12) V11
(13) V12
(14) V13
(15) V14
(16) V15
(17) V16
(18) V17
(19) V18
(20) V19
(21) V20
(22) V21
(23) V22
(24) V23
(25) V24
(26) V25
(27) V26
(28) V27
(29) V28
(30) Amount
(31) Fraud Class (Target- Dependent Attribute)

B. Prediction of Fraud Credit Card Transaction

Credit Card Transaction Data Set is implemented to analyze the target distribution of fraud class and is shown in Fig. 2 and Fig. 3.

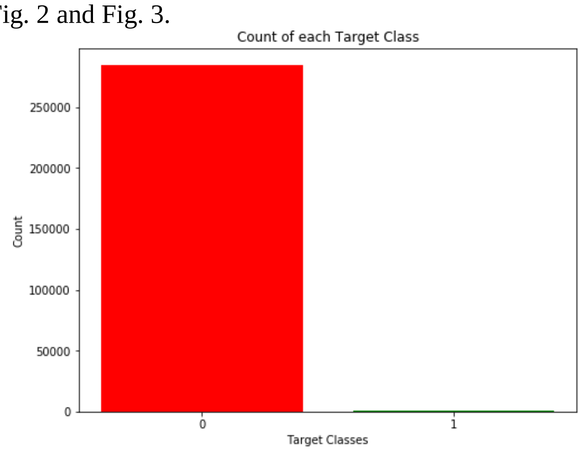

Fig. 2 Feature importance of attributes in the dataset

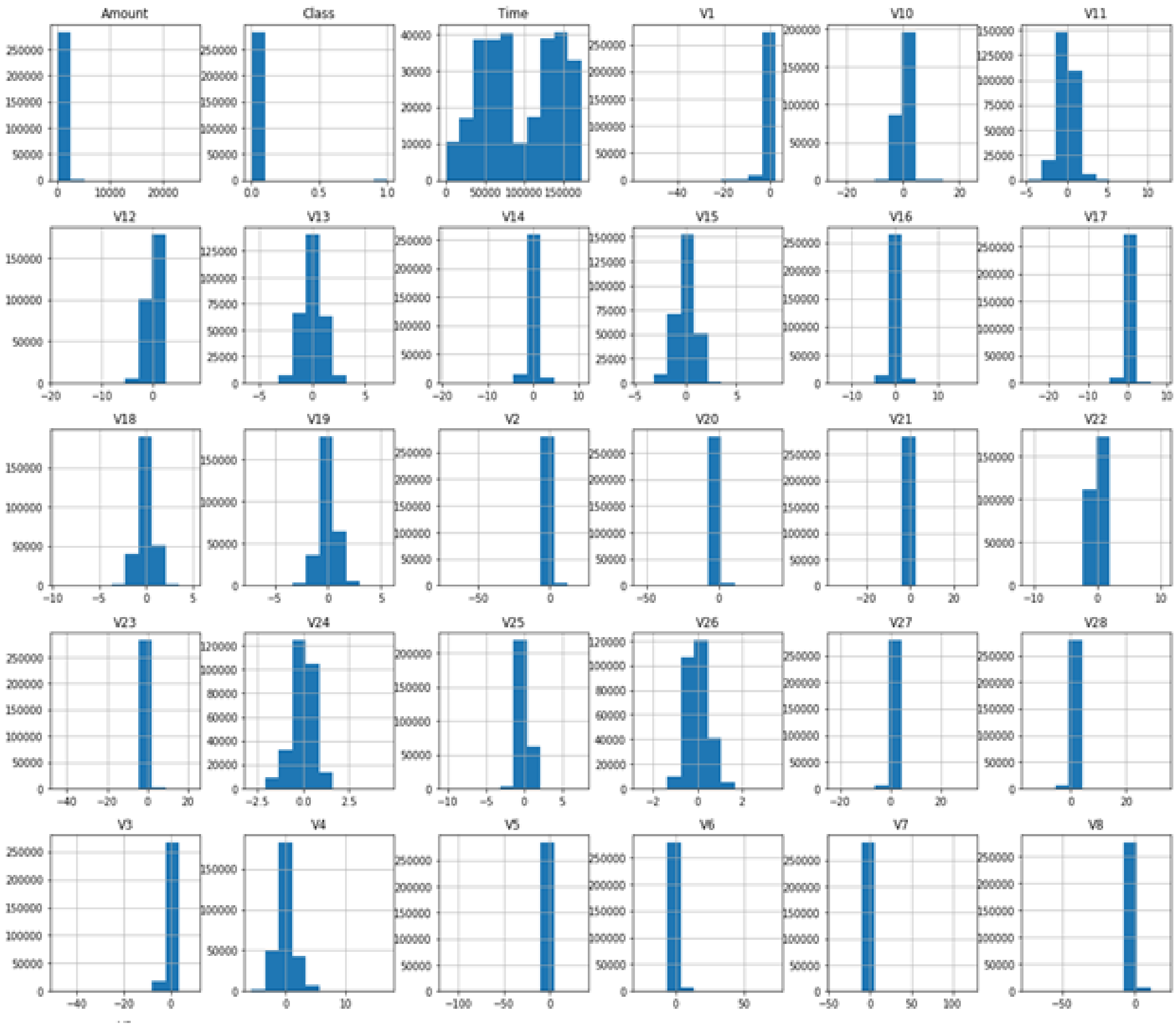

Fig. 3 Histogram distribution of credit card transaction Dataset

The Credit Card Transaction Data Set class target distribution is shown in Fig. 4.

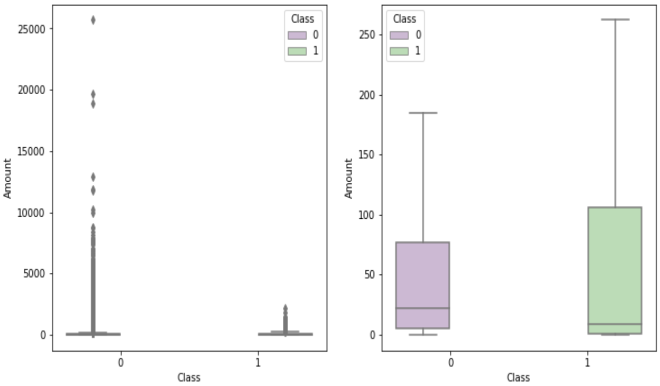

Fig. 4. Class distribution in the Data Set

The correlation between the variables of the Credit Card Transaction Data Set is shown in Fig. 5.

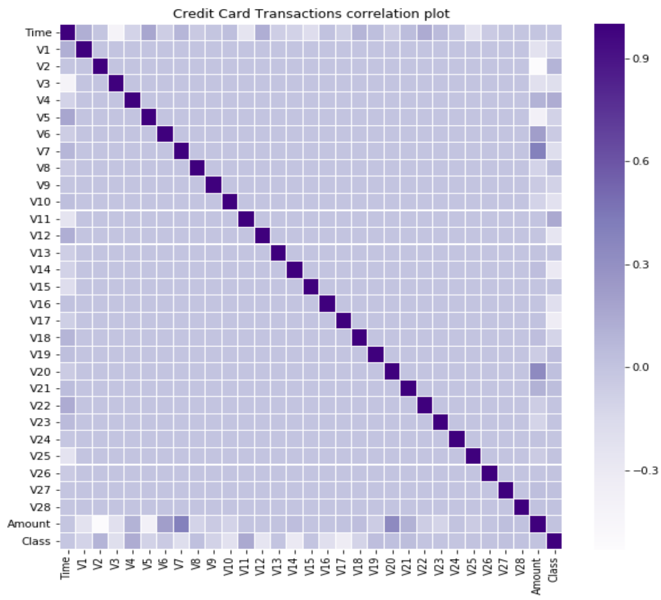

Fig. 5 Correlation matrix of Credit card transaction Data Set

The relationship between the transaction amount variable with the top features of the Credit Card Transaction Data Set shown in Fig. 6.

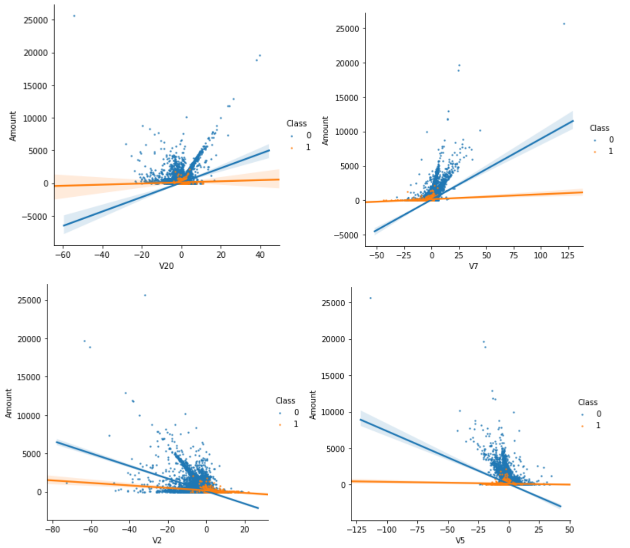

Fig. 6. Amount VS top features in the Data Set

The histogram relationship of all the attributes in the Credit Card Transaction Data Set shown in Fig. 7.

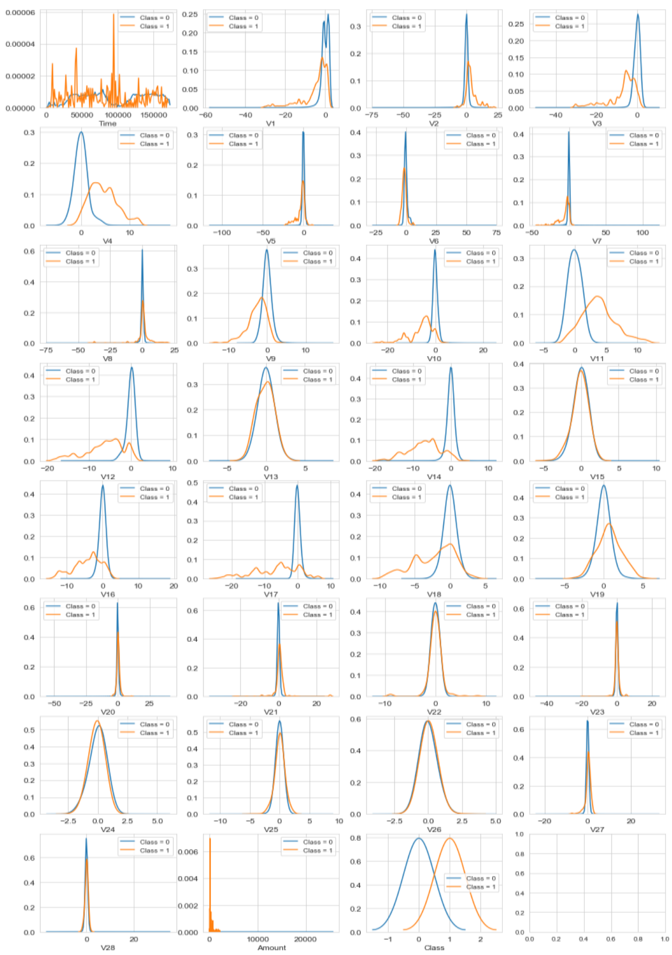

Fig. 7. Histogram relationship of the Data Set

The feature importance of the dataset is identified using Random Forest classifier and its feature importance along with the obtained confusion matrix is shown in Fig 8- Fig. 9.

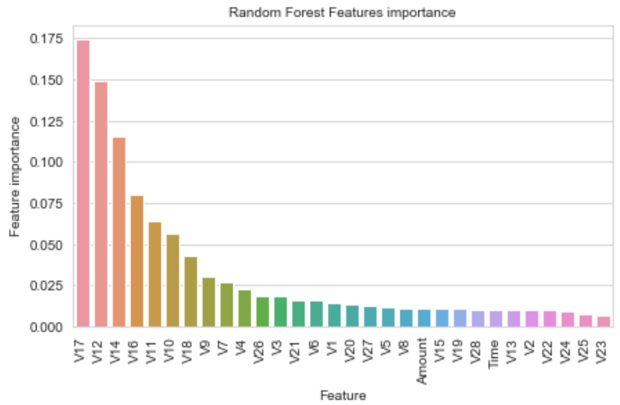

Fig. 8. Random Forest Feature Importance of the Data Set

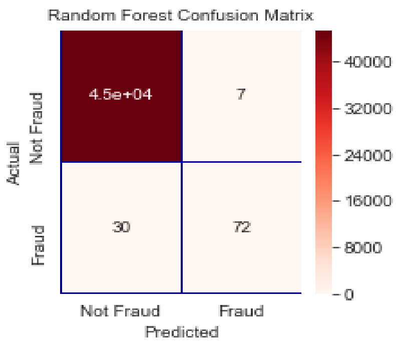

Fig. 9. Random Forest Confusion Matrix

The feature importance of the dataset is identified using Ada Boost classifier and its feature importance along with the obtained confusion matrix is shown in Fig 10- Fig. 11.

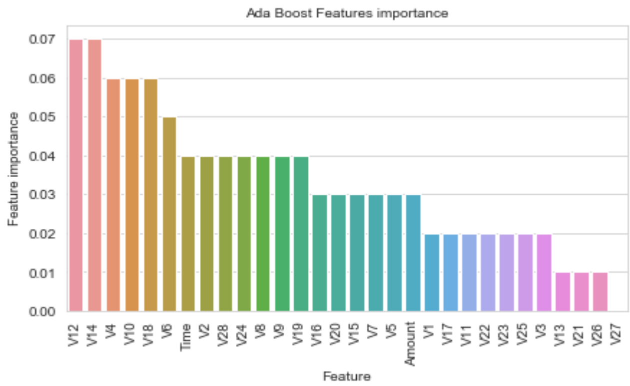

Fig. 10. Ada Boost Feature Importance

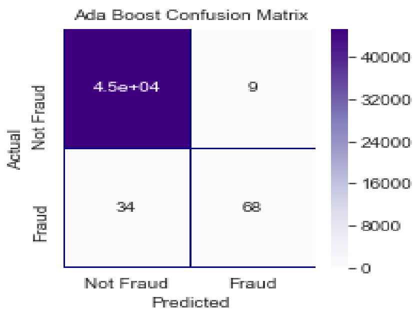

Fig. 11. Ada Boost Confusion Matrix

The confusion matrix of the dataset is identified using logistic regression classifier is shown in Fig 12.

Retrieval Number: D7621118419/2019©BEIESP
DOI:10.35940/ijrte.D7621.118419
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

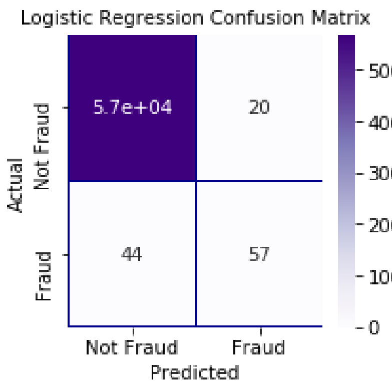

Fig. 12. Logistic Regression Confusion Matrix

The feature importance of the dataset is identified using Extra Tree classifier and its feature importance along with the obtained confusion matrix is shown in Fig 13- Fig. 14.

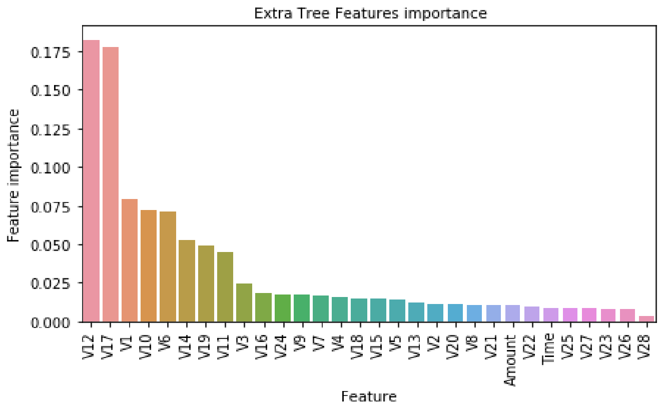

Fig. 13. Extra Tree Feature Importance of the Data Set

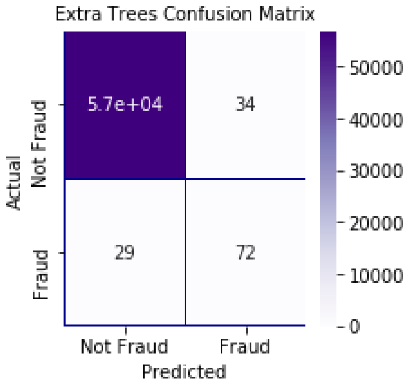

Fig. 14. Extra Tree Confusion Matrix

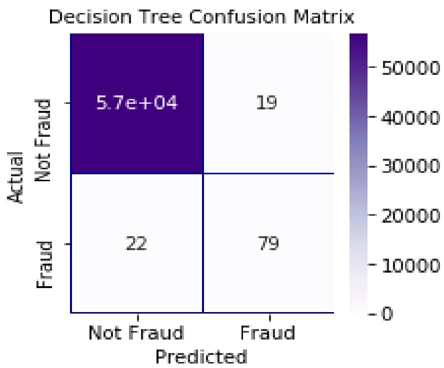

Fig. 15. Decision Tree Confusion Matrix

The feature importance of the dataset is identified using Decision Tree classifier and its feature importance along with the obtained confusion matrix is shown in Fig 15- Fig. 16.

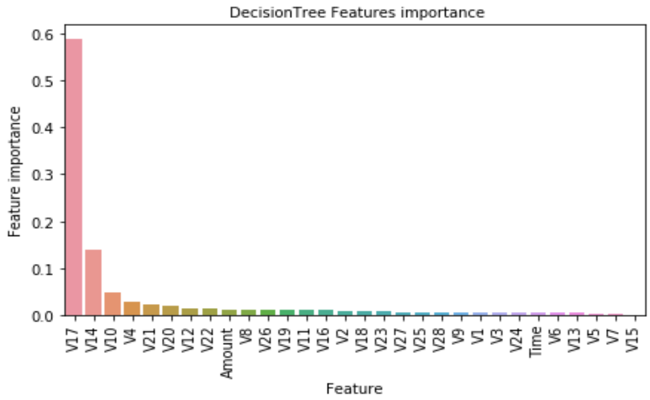

Fig. 16. Decision Tree Feature Importance of the Data Set

The feature importance of the dataset is identified using Naïve Bayes and Gradient Boosting classifier and its feature importance along with the obtained confusion matrix is shown in the Fig 17 – Fig 19.

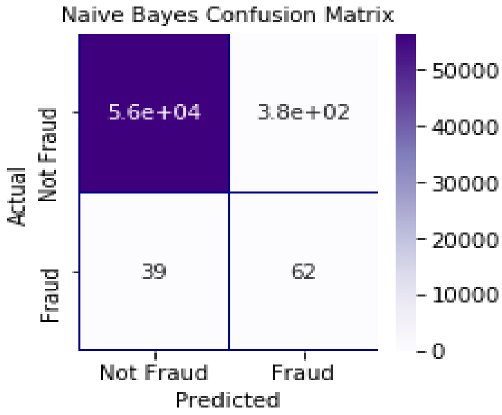

Fig. 17. Naïve Bayes Confusion Matrix

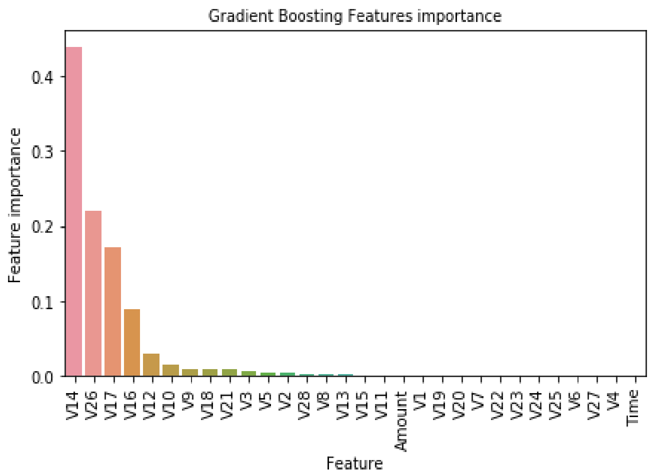

Fig. 18. Gradient Boosting Feature Importance of the Data Set

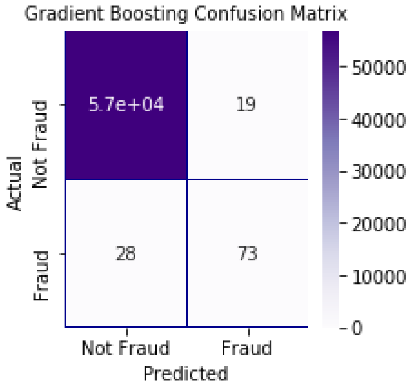

Fig. 19. Gradient Boosting Confusion Matrix

Performance analysis is done by analyzing the performance metrics like Accuracy, FScore, Precision and Recall is shown in Table 1 – Table 3.

Table. 1. Estimation of Classifier Parameters

Classifier Methods	Precision	Recall
Random Forest classifier	1.0	1.0
Ada Boost classifier	1.0	1.0
Logistic Regression	1.0	1.0
Decision Tree classifier	1.0	1.0
Extra Tree classifier	1.0	1.0
Gradient Boosting	1.0	1.0
Naive Bayes classifier	1.0	0.99

Table. 2. AUC Score Estimation of Classifier Parameters

Classifier Methods	FScore	AUC Score
Random Forest classifier	1.0	85.28
Ada Boost classifier	1.0	83.32
Logistic Regression	1.0	78.20
Decision Tree classifier	1.0	89.09
Extra Tree classifier	1.0	85.61
Gradient Boosting	1.0	86.12
Naive Bayes classifier	0.99	80.35

Table. 3. Accuracy Estimation of Classifier

Classifier Methods	Accuracy (%)
Random Forest classifier	99.91
Ada Boost classifier	99.90
Logistic Regression classifier	99.88
Decision Tree classifier	99.92
Extra Tree classifier	99.88
Gradient Boosting classifier	99.91
Naive Bayes classifier	99.21

The Performance Analysis of the metrics for all the regression is shown in Fig. 3. - Fig. 7.

V. CONCLUSION

This paper attempts to predict the fraud credit card transaction by extracting the credit card transaction dataset from the KAGGLE machine learning repository. The prediction of fraud transaction is done by using machine learning regression algorithms. The correlation between each of the dataset variables. Then the top most high correlated features are directly fitted to classification methods. Experimental Results shows that the Decision Tree classifier have achieved the effective prediction with the precision of 1.0, recall of 1.0, FScore of 1.0 , AUC Score of 89.09 and Accuracy of 99.92%..

REFERENCES

1. Joseph Janison, "Applying Machine Learning to Predict Davidson College's Admissions Yield", proceedings of the ACM SIGCSE Technical Symposium., 2017.
2. William Eberle, Douglas Talbert, Erik Simpson, Larry Roberts, and Alexis pope, " Using Machine Learning and Predictive Modeling to Assess Admission Policies and Standards", Proceedings of the 9th Annual National Symposium., 2013.
3. M. Shyamala Devi, Shakila Basheer, Rincy Merlin Mathew, "Exploration of Multiple Linear Regression with Ensembling Schemes for Roof Fall Assessment using Machine Learning", International Journal of Innovative Technology and Exploring Engineering, vol. 8, no.12, October 2019.
4. Shakila Basheer, Rincy Merlin Mathew, M. Shyamala Devi, "Ensembling Coalesce of Logistic Regression Classifier for Heart Disease Prediction using Machine Learning", International Journal of Innovative Technology and Exploring Engineering, vol. 8, no.12, October 2019, pp. 127-133.
5. Rincy Merlin Mathew, M. Shyamala Devi, Shakila Basheer," Exploration of Neighbor Kernels and Feature Estimators for Heart Disease Prediction using Machine Learning", International Journal of Innovative Technology and Exploring Engineering, vol. 8, no.12, October 2019, pp. 597-605.
6. M. Shyamala Devi, Shefali Dewangan, Satwat Kumar Ambashta, Anjali Jaiswal, Nariboyena Vijaya Sai Ram, "Backward Eliminated Formulation of Fire Area Coverage using Machine Learning Regression", International Journal of Innovative Technology and Exploring Engineering, vol. 8, no.12, October 2019, pp.1565-1569
7. M. Shyamala Devi, Ankita Shil, Prakhar Katyayan, Tanmay Surana, "Constituent Depletion and Divination of Hypothyroid Prevalance using Machine Learning Classification", International Journal of Innovative Technology and Exploring Engineering, vol. 8, no.12, October 2019, pp. 1607-1612
8. M. Shyamala Devi, Shefali Dewangan, Satwat Kumar Ambashta, Anjali Jaiswal, Sairam Kondapalli, "Recognition of Forest Fire Spruce Type Tagging using Machine Learning Classification", International Journal of Recent Technology and Engineering, Volume-8 Issue-3, pp. 4309 – 4313, 16 September 2019.
9. M. Shyamala Devi, Usha Vudatha, Sukriti Mukherjee, Bhavya Reddy Donthiri, S B Adhiyan, Nallareddy Jishnu, " Linear Attribute Projection and Performance Assessment for Signifying the Absenteeism at Work using Machine Learning", International Journal of Recent Technology and Engineering, Volume-8 Issue-3, pp. 1262 – 1267, 16 September 2019.
10. M. Shyamala Devi, Mothe Sunil Goud, G. Sai Teja, MallyPally Sai Bharath, "Heart Disease Prediction and Performance Assessment through Attribute Element Diminution using Machine Learning", International Journal of Innovative Technology and Exploring Engineering, vol. 8, no.11, pp. 604 – 609, 30 September 2019.
11. M. Shyamala Devi, Rincy Merlin Mathew, R. Suguna, " Regressor Fitting of Feature Importance for Customer Segment Prediction with Ensembling Schemes using Machine Learning", International Journal of Engineering and Advanced Technology, Volume-8 Issue-6, pp. 952 – 956, 30 August 2019.
12. R. Suguna, M. Shyamala Devi, Rincy Merlin Mathew, "Integrating Ensembling Schemes with Classification for Customer Group Prediction using Machine Learning", International Journal of Engineering and Advanced Technology, Volume-8 Issue-6, pp. 957 – 961, 30 August 2019.
13. Rincy Merlin Mathew, R. Suguna, M. Shyamala Devi, "Composite Model Fabrication of Classification with Transformed Target Regressor for Customer Segmentation using Machine Learning", International Journal of Engineering and Advanced Technology, Volume-8 Issue-6, pp. 962 – 966, 30 August 2019.

14. M. Shyamala Devi, Rincy Merlin Mathew, R. Suguna, "Feature Snatching and Performance Analysis for Connoting the Admittance Likelihood of student using Principal Component Analysis", International Journal of Recent Technology and Engineering, Volume-8 Issue-2, 30 July 2019. pp. 4800-4807.
15. R. Suguna, M. Shyamala Devi, Rincy Merlin Mathew, "Customer Segment Prognostic System by Machine Learning using Principal Component and Linear Discriminant Analysis", International Journal of Recent Technology and Engineering, Volume-8 Issue-2, 30 July 2019. pp. 6198-6203.
16. R.Suguna, M. Shyamala Devi, Rupali Amit Bagate, Aparna Shashikant Joshi, "Assessment of Feature Selection for Student Academic Performance through Machine Learning Classification", Journal of Statistics and Management Systems, Taylor Francis, , vol. 22, no. 4, 25 June 2019, pp. 729-739. DOI: 10.1080/09720510.2019.1609729ISSN: 0972-0510 (Print), 2169-0014 (Online).
17. R.Suguna, M. Shyamala Devi, Rupali Amit Bagate, Aparna Shashikant Joshi, "Assessment of Feature Selection for Student Academic Performance through Machine Learning Classification", Journal of Statistics and Management Systems, Taylor Francis, vol. 22, no. 4, 25 June 2019, pp. 729-739. DOI: 10.1080/09720510.2019.1609729ISSN: 0972-0510 (Print), 2169-0014 (Online).
18. Shyamala Devi Munisamy, Suguna Ramadass Aparna Joshi, "Cultivar Prediction of Target Consumer Class using Feature Selection with Machine Learning Classification", Learning and Analytics in Intelligent Systems, LAIS, Springer, vol. 3, pp. 604-612, June 2019.
19. Suguna Ramadass, Shyamala Devi Munisamy, Praveen Kumar P, Naresh P, "Prediction of Customer Attrition using Feature Extraction Techniques and its Performance Assessment through dissimilar Classifiers", Springer's book series entitled "Learning and Analytics in Intelligent Systems, Springer, LAIS vol. 3, pp. 613-620, June 2019.
20. M. Shyamala Devi, Rincy Merlin Mathew, R. Suguna, "Attribute Heaving Extraction and Performance Analysis for the Prophesy of Roof Fall Rate using Principal Component Analysis", International Journal of Innovative Technology and Exploring Engineering, vol. 8, no.8, June 2019, pp. 2319-2323.
21. R. Suguna, M. Shyamala Devi, Rincy Merlin Mathew, "Customer Churn Predictive Analysis by Component Minimization using Machine Learning", International Journal of Innovative Technology and Exploring Engineering, vol. 8, no.8, June 2019, pp. 2329-2333.