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 has achieved the effective prediction with the precision of 1.0, recall of 1.0, FScore of 1.0, AUC Score of 0.99 and Accuracy of 99.92%.

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

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 analysis 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 stolen 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.
Swindling Shonky Anatomization of Credit Card Transactions using Machine Learning

(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

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

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

Fig. 3 Histogram distribution of credit card transaction Dataset

Fig. 4. Class distribution in the Data Set

Fig. 5 Correlation matrix of Credit card transaction Data Set
Swindling Shonky Anatomization of Credit Card Transactions using Machine Learning

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](image)

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](image)

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](image)

![Fig. 9. Random Forest Confusion Matrix](image)

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](image)

![Fig. 11. Ada Boost Confusion Matrix](image)

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

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.

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.
Swindling Shonky Anatomization of Credit Card Transactions using Machine Learning

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**

<table>
<thead>
<tr>
<th>Classifier Methods</th>
<th>Precision</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random Forest classifier</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Ada Boost classifier</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Logistic Regression</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Decision Tree classifier</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Extra Tree classifier</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Gradient Boosting</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Naive Bayes classifier</td>
<td>0.99</td>
<td></td>
</tr>
</tbody>
</table>

**Table. 2. AUC Score Estimation of Classifier Parameters**

<table>
<thead>
<tr>
<th>Classifier Methods</th>
<th>FScore</th>
<th>AUC Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random Forest classifier</td>
<td>1.0</td>
<td>85.28</td>
</tr>
<tr>
<td>Ada Boost classifier</td>
<td>1.0</td>
<td>83.32</td>
</tr>
<tr>
<td>Logistic Regression</td>
<td>1.0</td>
<td>78.20</td>
</tr>
<tr>
<td>Decision Tree classifier</td>
<td>1.0</td>
<td>89.09</td>
</tr>
<tr>
<td>Extra Tree classifier</td>
<td>1.0</td>
<td>85.61</td>
</tr>
<tr>
<td>Gradient Boosting</td>
<td>1.0</td>
<td>86.12</td>
</tr>
<tr>
<td>Naive Bayes classifier</td>
<td>0.99</td>
<td>80.35</td>
</tr>
</tbody>
</table>

**Table. 3. Accuracy Estimation of Classifier**

<table>
<thead>
<tr>
<th>Classifier Methods</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random Forest classifier</td>
<td>99.91</td>
</tr>
<tr>
<td>Ada Boost classifier</td>
<td>99.90</td>
</tr>
<tr>
<td>Logistic Regression classifier</td>
<td>99.88</td>
</tr>
<tr>
<td>Decision Tree classifier</td>
<td>99.92</td>
</tr>
<tr>
<td>Extra Tree classifier</td>
<td>99.88</td>
</tr>
<tr>
<td>Gradient Boosting classifier</td>
<td>99.91</td>
</tr>
<tr>
<td>Naive Bayes classifier</td>
<td>99.21</td>
</tr>
</tbody>
</table>

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**


