

Capstone Project

Project Title: **Predict the Creditcard Fraud in Given Dataset**

Abstract:

The goal of this project is to analyse the given bank data and predict the whether a People make a fraud or not. The data was divided into two parts i.e. the training and the test dataset. Models were developed based on the training dataset and applied to the test dataset to find out the accuracy of each model based on the predicted values generated. Based on these values we can determine how good a particular model is for prediction of credit card fraud

Submitted By:
Ajith Ramesh

Table of Content

Introduction	2
Feature Explanation	3
Scope	4
Steps performed.....	5
Results.....	6
Conclusion.....	7

Predict the credit card fraud in transaction

Introduction

In today's modern era of electronic payments and online shopping, credit cards have become an indispensable tool for financial transactions. However, amidst the convenience and ease they offer, there exists a lurking danger known as credit card fraud. This illicit practice involves unauthorized access to someone's credit card information, leading to fraudulent charges and potential financial loss

Capstone Project

Dataset Information

According to the Data Set has more than 5 million records been being stolen on a daily basis, a concerning statistic that shows - fraud is still very common both for Card-Present and Card-not Present type of payments. The Digital payments are evolving, but so are cyber criminals. we experiment with the credit card fraud dataset to explore the machine learning algorithms and build an optimum model to predict the fraud.

Feature Explanation:

Column	Description
distance_from_home	the distance from home where the transaction happened.
distance_from_last_transaction	the distance from last transaction happened.
ratio_to_median_purchase_price	Ratio of purchased price to transaction median purchase price.
repeat_retailer	Is the transaction happened from same retailer.
used_chip	Is the transaction through chip (credit card).
used_pin_number	Is the transaction happened by using PIN number.
online_order	Is the transaction an online order.
fraud	Is the transaction fraudulent.

Capstone Project

Scope :

- Analysing the available data and exploring relationships among given variables
- Data Pre-processing
- Visualizing the data in the dataset
- Build the classification model using Logistic Regression, Decision Tree, AdaBoostClassifier, GaussianNB and KNeighborsClassifier

Steps Performed

1. Installing the necessary packages.
2. Fetching the required packages from the library.
3. Cleaning the data.
4. Developing various plots for exploratory analysis.
5. Creating the correlation
5. Creating a cleaner dataset for analysis
6. Creating a Training and Test dataset with 60% being the training data and 40% being test data

7.Model Building

In the dataset we build the five supervised classification model to predict the best accuracy and also using the future prediction

- a. Model 1- Using Logistic Regression
- b. Model 2 – Using Decision Tree
- c. Model 3 – Ensemble model using AdaboostClassifier
- d. Model 4 – Naive Bayes model using GaussianNB
- e. Model 5 – Using KNeighbors Classifier

7. Predicting Values

8. Creating the actual confusion matrix based on predicted values.

9. Tabulate the Result

Result

	Model	AUC score	Precision Score	Recall Score	Accuracy Score	Kappa Score	f1 Score
0	Logistic Regression	0.966621	0.889904	0.607144	0.959255	0.700682	0.721820
1	Decision_tree_model	0.999912	0.999971	0.999828	0.999982	0.999890	0.999899
2	AdaBoostClassifier	0.999995	0.998221	0.999139	0.999770	0.998554	0.998680
3	GaussianNB	0.964009	0.787609	0.591696	0.950558	0.649561	0.675739
4	KNeighborsClassifier	0.994669	0.881408	0.919775	0.982240	0.890440	0.900183

Conclusion

The supervised classification learning algorithms named in the above table have been implemented on the given dataset. The performance of the models were evaluated using AUC score, precision, recall, accuracy, kappa score and f1-score.

To conclude I would like to say Decision Tree model has the highest values for most of the performance measures like AUC Score, Recall, f1-score, accuracy. Therefore, it can be concluded that the Decision Tree model can be used to predict the existence of creditcard fraud prediction. In future when we have more fraud detection or bigger dataset and we can also apply Decision Tree model and AdaBoostClassifier and ensemble techniques for prediction that can be tested again.