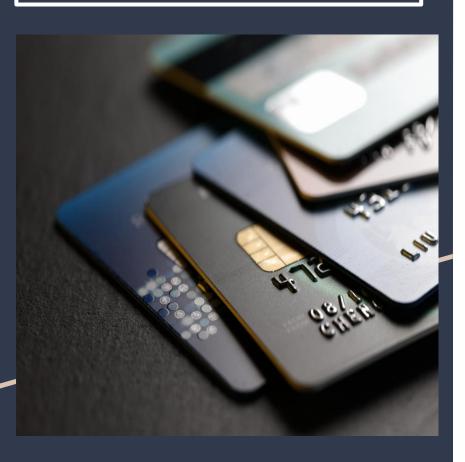# Credit Card Fraud Detection

CASE STUDY
MACHINE LEARNING
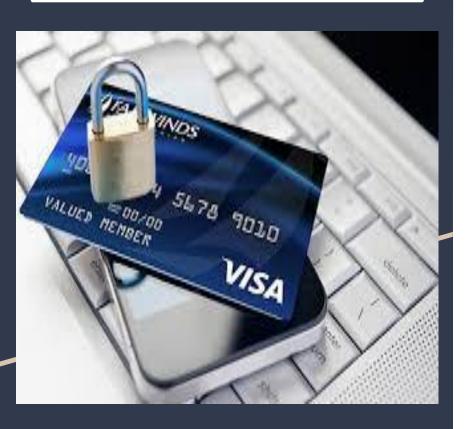
NAME :- ANJALI JHA

INTRODUCTION:-

- **Fraud detection refers to the detection of criminal activities occuring in commercial organisations.**

- **In other words we can say credit card fraud detection is when businesses help us prevent stolen money , products and services obtained through illegitimate credit transactions.**

- **Fraud can happen by the cardholder or even possible by someone else(identity theft).**

OBJECTIVES/NEED:-

● Fraud can cause major losses to individuals or companies.

● The cost of a fraud prevention program is tiny compared to a major failure.

● The process of developing a fraud prevention program/model is beneficial both because it will help prevent future frauds and also because it is a process of discovery

● Efforts to reduce the risk of fraud makes you a better investment, business partner and supplier.

PROBLEM STATEMENT :-

The Credit Card Fraud Detection includes considering /modelling past credit card transaction with the knowledge of the ones that turned out to be fraud.

This model is then used to identify whether a new transaction is fraudulent or not. Our aim is to detect 100% of the fraudulent transactions while minimising the misclassifications i.e The incorrect fraud classification.

CHALLENGES:-

1. To recognise fraudulent credit card transactions so that the customers of credit card

companies are not charged for the items that they did not purchase.

2. Unbalanced data i.e most of the transactions are not fraud which makes it really hard to

detect the fraudulent ones.

3. Misclassification cost is very high.

4. Misclassification Data can be another major issue as not every fraudulent transaction is

reported or caught.

5. Adaptive technique used against the models by the scammer

DESCRIPTION:-

1. Dataset is in csv format.

2. Dataset contains 284807 and 31 columns.

3. There are no missing values present in the dataset.

4. The output of the dataset is 'Class' and This notifies whether the transaction is fraud or not.

EDA AND PREPROCESSING:-

Exploratory Data Analysis(EDA)is used for cleaning the data

● **Handle missing values**

● **Handle outliers**

● **Handle skewness**

● **Handling categorical data**

● **Feature Engineering**

After cleaning the data we try separating the input and the output in X and Y variable.

● **Take 70-30 pattern for training and testing process.**
● **Sampling is applied for balancing the data if data is imbalance. Random oversampling is used mainly**
● **ROS is used to increase the dataof "yes" or fraud"1" as the available data of yes class is very less in comparison to no class.**
● **Scaling on training and testing data**

METHODS/ALGO USED:-

1. **Logistic Regression**

2. **Decision Tree Classifier**

3. **Ensembling Technique:**

- **.Naive Aggregation**
- **Bootstrapping**

    a)Bagging

    b)Pasting

    c)Random Forest

- **Boosting**
- **Stacking Classifier**

5.**Support Vector Machine (SVM)**

Logistic Regression:-Its the process of converting the input variable into expression of probability that ranges between 0 and 1.

```
In [18]: lr=create_model(lr)
```

	precision	recall	f1-score	support
0	1.00	0.98	0.99	85308
1	0.07	0.90	0.13	135
accuracy			0.98	85443
macro avg	0.54	0.94	0.56	85443
weighted avg	1.00	0.98	0.99	85443

```
confusion matrix
[[83713   1595]
 [   14    121]]
```

Decision Tree Classifier

```
In [26]: dt1=create_model(dt1)
```

	precision	recall	f1-score	support
0	1.00	0.99	1.00	85308
1	0.16	0.83	0.27	135
accuracy			0.99	85443
macro avg	0.58	0.91	0.63	85443
weighted avg	1.00	0.99	1.00	85443

```
confusion matrix
[[84721   587]
 [   23   112]]
```

```
[33]: dt3=create_model(dt3)
```

	precision	recall	f1-score	support
0	1.00	1.00	1.00	85308
1	0.35	0.80	0.48	135
accuracy			1.00	85443
macro avg	0.67	0.90	0.74	85443
weighted avg	1.00	1.00	1.00	85443

```
confusion matrix
[[85105   203]
 [   27   108]]
```

ENSEMBLING TECHNIQUE(Bagging)

```
[44]:  bc=create_model(bc)
```

	precision	recall	f1-score	support
0	1.00	0.98	0.99	85308
1	0.07	0.90	0.13	135
accuracy			0.98	85443
macro avg	0.54	0.94	0.56	85443
weighted avg	1.00	0.98	0.99	85443

```
confusion matrix
[[83714   1594]
 [   14    121]]
```

CONCLUSION:-

Our ultimate goal of this project is to determine the algorithm for analysis as well as the best performing combination of factors which help us in detecting credit card fraud.

RESULT:-

Achieved quite good results by logistic Regression and also by bagging model of ensembling technique in this particular dataset of credit card.