

# CREDIT CARD FRAUD DETECTION



# Introduction of Project

- > Credit card fraud detection dataset consist fraudulent and fair list of transaction which helps us to identify or classify the difference between fraudulent and fair transaction, from this identification or classification we can take preventing measures to stop the fraudulent transaction.
- > Fraud credit card transaction is a big issue for customers and as well as for banks because this leads to loss of money for a customer and may also lead to loss of customer for bank.
- > So to retain the customers bank has to make necessary steps to identify and stop such fraud transactions.

# PROJECT OBJECTIVE

To predict the fraud transaction.

Identifying different patterns or methods used in the fraud transactions.

Using different ML algorithms to predict the best output.

Finding the best model for fraud prediction and providing summary

# DATASET DESCRIPTION

- > Source dataset is in csv format
- > Dataset contains 284807 rows and 31 columns
- > There are no missing values in the provided in the dataset The output of the dataset is “Class” which notifies whether the transaction is fraud or fair

# **DATASET OVERVIEW AND PREPROCESSING**

- > After checking the data set it is seen that the data is clean and there are no null values.**
- > Random over sampler is also used to increase the data of “Yes or Fraud (1)” Class because the available data of Yes class was very less compared to No class.**

# ALGORITHMS –

- 1.** To solve this classification dataset we have used below algorithms after dividing in training and testing dataset  
Logistic Regression
- 2.** Decision Tree Classifier (using Gini index, entropy method and random forest)
- 3.** Support Vector Machine

# ENSEMBLING TECHNIQUES

## 1. Bootstrapping

- Random forest tree

## 2. Boosting Technique

- Ada boosting
- Gradient Boosting

## 3. Stacking

- Extreme gradient Boosting

# CLASSIFICATION REPORT AND CONFUSION MATRIX OF LOGISTIC REGRESSION

	precision	recall	f1-score	support
0	0.88	0.97	0.92	85308
1	0.97	0.87	0.91	85308
accuracy			0.92	170616
macro avg	0.92	0.92	0.92	170616
weighted avg	0.92	0.92	0.92	170616

```
[[82827 2481]
 [11452 73856]]
```

# CLASSIFICATION REPORT AND CONFUSION MATRIX OF DECISION TREE (GINI INDEX, ENTROPY)

## GINI INDEX

	precision	recall	f1-score	support
0	0.76	1.00	0.86	85308
1	1.00	0.68	0.81	85308
accuracy			0.84	170616
macro avg	0.88	0.84	0.84	170616
weighted avg	0.88	0.84	0.84	170616

Confusution matrix  
[[85280 28]  
[27256 58052]]

## ENTROPY

	precision	recall	f1-score	support
0	0.83	1.00	0.91	85308
1	1.00	0.80	0.89	85308
accuracy				0.90
macro avg	0.91	0.90	0.90	170616
weighted avg	0.91	0.90	0.90	170616

Confusution matrix  
[[85104 204]  
[17090 68218]]

# CLASSIFICATION REPORT AND CONFUSION MATRIX OF DECISION TREE (PURNING TECHNIQUE)

## Max depth

	precision	recall	f1-score	support
0	0.84	1.00	0.91	85308
1	0.99	0.81	0.90	85308
accuracy			0.90	170616
macro avg	0.92	0.90	0.90	170616
weighted avg	0.92	0.90	0.90	170616

Confusution matrix  
[[84900 408]  
[15820 69488]]

## Minimum Sample leaf

	precision	recall	f1-score	support
0	0.83	1.00	0.91	85308
1	1.00	0.79	0.88	85308
accuracy			0.90	170616
macro avg	0.91	0.90	0.89	170616
weighted avg	0.91	0.90	0.89	170616

Confusution matrix  
[[85155 153]  
[17707 67601]]

# Best Classification Report and Confusion Matrix Found From All Ensembling Technique

	precision	recall	f1-score	support
0	0.90	0.97	0.93	85308
1	0.97	0.89	0.93	85308
accuracy			0.93	170616
macro avg	0.93	0.93	0.93	170616
weighted avg	0.93	0.93	0.93	170616

  

```
[[82833 2475]
 [ 9502 75806]]
```



# CONCLUSION

- After working over the dataset we have predicted the fraud and fair transactions, so now we can actually take preventing measures to identify and block such fraud transaction in future.
- This prediction will also help us in identifying the different patterns or ways of executing the fraud transaction.
- This will help us to reduce the fraud transaction and so the people can use their credit card unworriedly.
- There might be some transaction in which the machine may get confused in concluding the fraud and fair transaction because machine learning does not give 100% correct prediction and therefore we have seen that after using several techniques and algorithms we have not seen any algorithm giving 100% prediction.
- The original data available for fraud transaction was very less compared to fair transaction therefore we surely need more original data to predict the different fraud transactions accurately.