KARTIK THAKUR.

Credit Card Frauds Detection (Machine learning)

* Kaggle dataset used :→

(1) import necessary lib → Sklearn, numpy, pandas, matplotlib, Seaborn (→ for co-relation), Scipy, math

Linear regression ✗ due to labels (Y/N) Classification problem.

decision tree
random forest classifier
Support Vector machine
confusion matrix
accuracy score
Standard Scaler
→ we can use these lib.

df → dataframe from github repository
df = pd.read_csv "——— ")

(2) reading dataset of Kaggle. (31 column) → label
class will tell its fraud or not.
(3) .head() [top 5 transcation] training data
(4) .info [about info of dataset]
(5) .iloc [:, 1:30].values → for location.
(6) X.shape, Y.shape → Tells about the input output range
(7) .isnull [to check for null value for replacing, mean, median]
(8) Counting & plotting in histogram.
(9) fraud data range & normal data range.
(10) .describe () [Statistical Summary]
(11) Co-relation
(12) Test data split.
(13) Standard scaler [convert to proper format for scaling (standardized)]
(14) Print standardized data.
(15) Decision tree classifier (& check & merge proper label prediction)
Entropy → used for gaining information.
(16) Prediction → for outside data to be predicted or not.
(17) Confusion matrix (Predicted & normal value matrix)
creates a

```
[[71052   30]
 [   25   95]]
```

KARTIK THAKUR.

Total 71,202		Predicted	
		NO	YES
ACTUAL	NO	T.N 71052 (0,0)	F.P 30 (0,1)
	YES	F.N 25 (1,0)	T.P 95 (1,1)
		71077	125

71082

120

① Accuracy = $\frac{TP+TN}{T}$

$= \frac{71052+95}{71202} = 99.92275$

② Error rate : 1 − accuracy

$1 - 0.999227$

$= 0.077245$

OR $\frac{FP+FN}{T} = 0.077245$

③ Precision-Specificity = $\frac{TP}{\text{Predicted Yes}} = \frac{95}{125}$

$= 76.0$

④ Sensitivity = $\frac{TN}{\text{Pre. NO}} = \frac{71052}{71077} = 99.9648$

⑱ Kernel work from low dimension to high dimension.

Excel sheet → with csv format in github anyones repository.
[dataset for fraud detection]