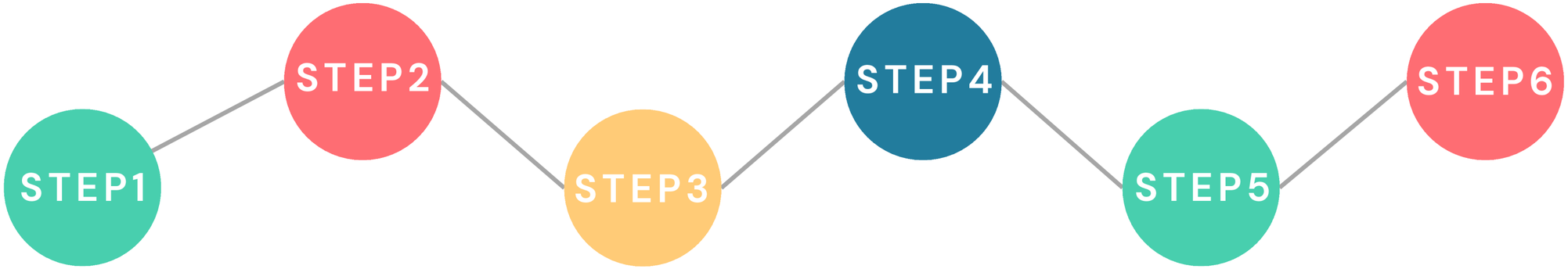
FRAUD DETECTION IN FINANCIAL TRANSACTIONS

Problem Description : Our mission is to create a model capable of detecting fraudulent activities in financial transactions. The challenge lies in accurately identifying irregular patterns and anomalies that signify fraudulent behavior while minimizing false positives.

DATASET LINK

Credit Card Transactions Fraud Detection Dataset (kaggle.com)

STEPS INVOLVED



Importing Python libraries

Loading Dataset in Jupyter NB

Data Preprocessing and Visualization

Feature Engineering and Label Encoding

Model Building

Model Evaluation

MODEL SELECTION APPROACH

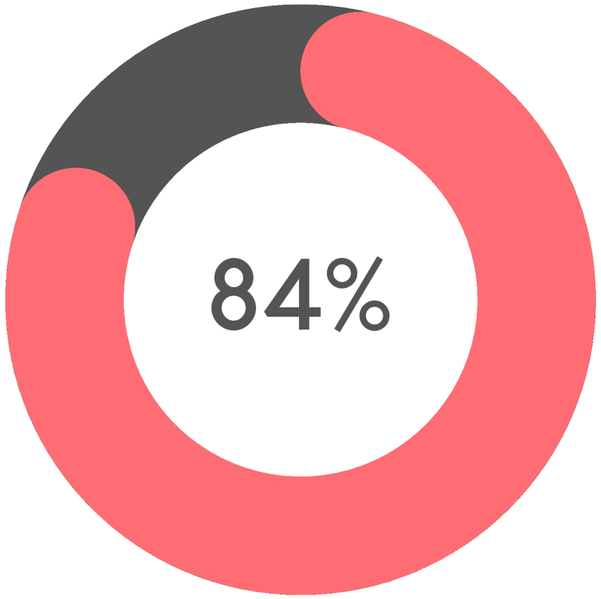
01 - LOGISTIC REGRESSION

02 - DECISION TREE

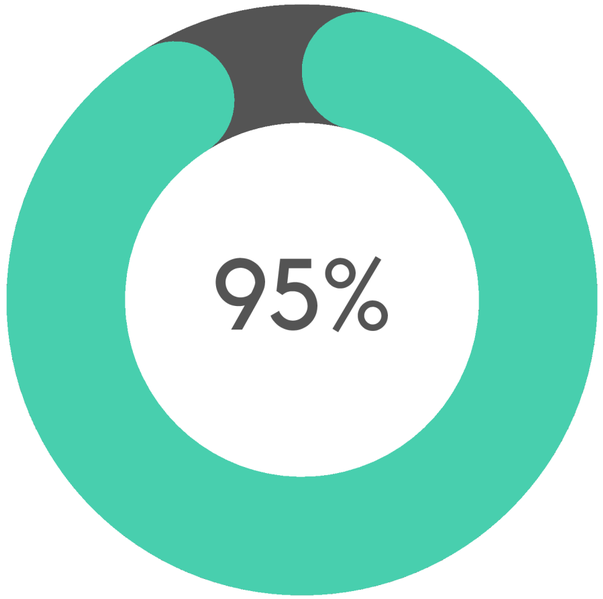
03 - SV MACHINE

As it is a Classification problem means we have to detect whether it is fraud or not .For this we have used these algorithms to implement

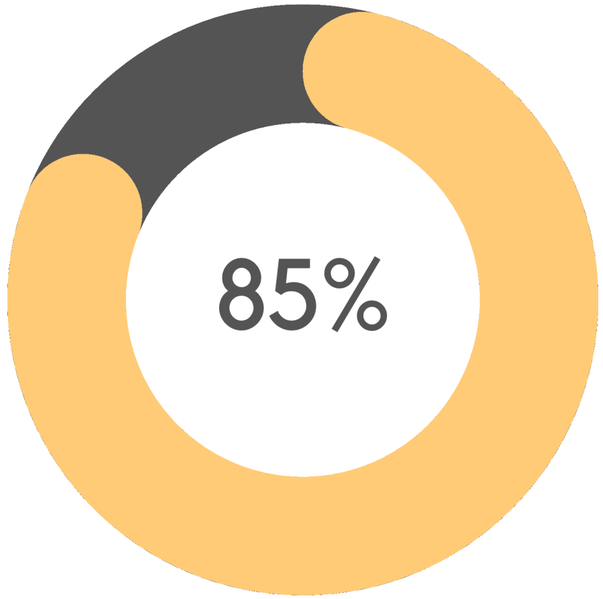
ACCURACY OF SELECTED MODELS



LOGISTIC REGRESSION



DECISION TREE



SV MACHINE

CHALLENGES OCCURED

1.Handling Large Dataset Challenge : In the credit card fraud detection project, we faced a significant challenge due to the large size of the dataset. Addressing this challenge was crucial for ensuring efficient model development and accurate predictions.

Approach to Handle this Challenge :

We use data Sampling to manage the dataset size during this , we employed random sampling to create a smaller subset of the data .

.

2. Dealing with Categorical Columns : It is also one challenge to deal with many categorical columns in the dataset , Because Categorical columns with a large number of unique values can lead to sparse representations and increase the dimensionality of the dataset, making it more complex to process.

Approach to Handle this Challenge :
Techniques like Label Encoding or One hot Encoding applied on categorical columns with a manageable number of unique values.This method creates binary columns for each category, allowing the model to interpret the presence or absence of each category.

3.Memory Optimization Challenge : There are many columns present in the dataset which need memory for their storage ,that is also a challenge to optimize the memory ,the large dataset was problematic, leading to slow data retrieval and processing times.

Approach to Handle this Challenge :
We performed feature selection to identify and retain only the most relevant features for model training. By this only crucial features are taken from the dataset and rest of the features we have removed from the dataset which reduce the memory space or storage .

INSIGHTS OBTAINED

1.High Fraud Occurrence in Middle Age Groups: The age groups between 30 to 50 years old show the highest counts of fraud cases. These age ranges are often associated with individuals actively engaged in financial activities like loans, credit card usage, and investments, making them more susceptible targets.

2.Peak Fraud Cases Around Age 40: There is a noticeable peak around the age of 40, with the count of fraud cases nearing 200. This suggests that individuals in this age range may be more exposed to risks or more actively involved in financial transactions.

3.Lower Fraud Incidents Among Younger and Older Age Groups: The fewer fraud cases in younger age groups (below 20) and older age groups (above 60). This could indicate less financial activity or different spending and financial management behaviors in these demographics.

THANK YOU