CREDIT CARD FRAUD DETECTION

RAJIV GANDHI UNIVERSITY OF KNOWLEDGE AND TECHNOLOGY,RK VALLEY

Department of Electronics and Communication Engineering(ECE)

Under the guidance of
B MOHAN REDDY

BY
GALETI VARSHALATHA
R161213-E4

KEY CONCEPTS:

- Project Idea.
- Tools and Libraries used.
- Introduction to Credit Card fraud.
- Steps to develp the classifier in ML.
- Algorithm used.
- How Random Forest Algorithm works.
- Why use of Random Forest Algorithm.
- Confusion matrix.
- Important steps to avoid fraud.

Project Idea:

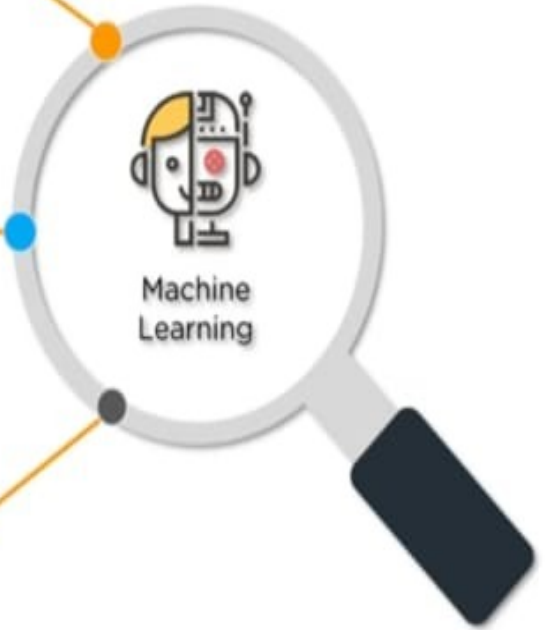
- The Objective of the Machine Learning Project is to detect the Credit Card Fraud Transactions using Python Libraries. We overcome the problem by creating a binary classifier and experimenting with machine learning technique.

Types of Machine Learning

Supervised Learning

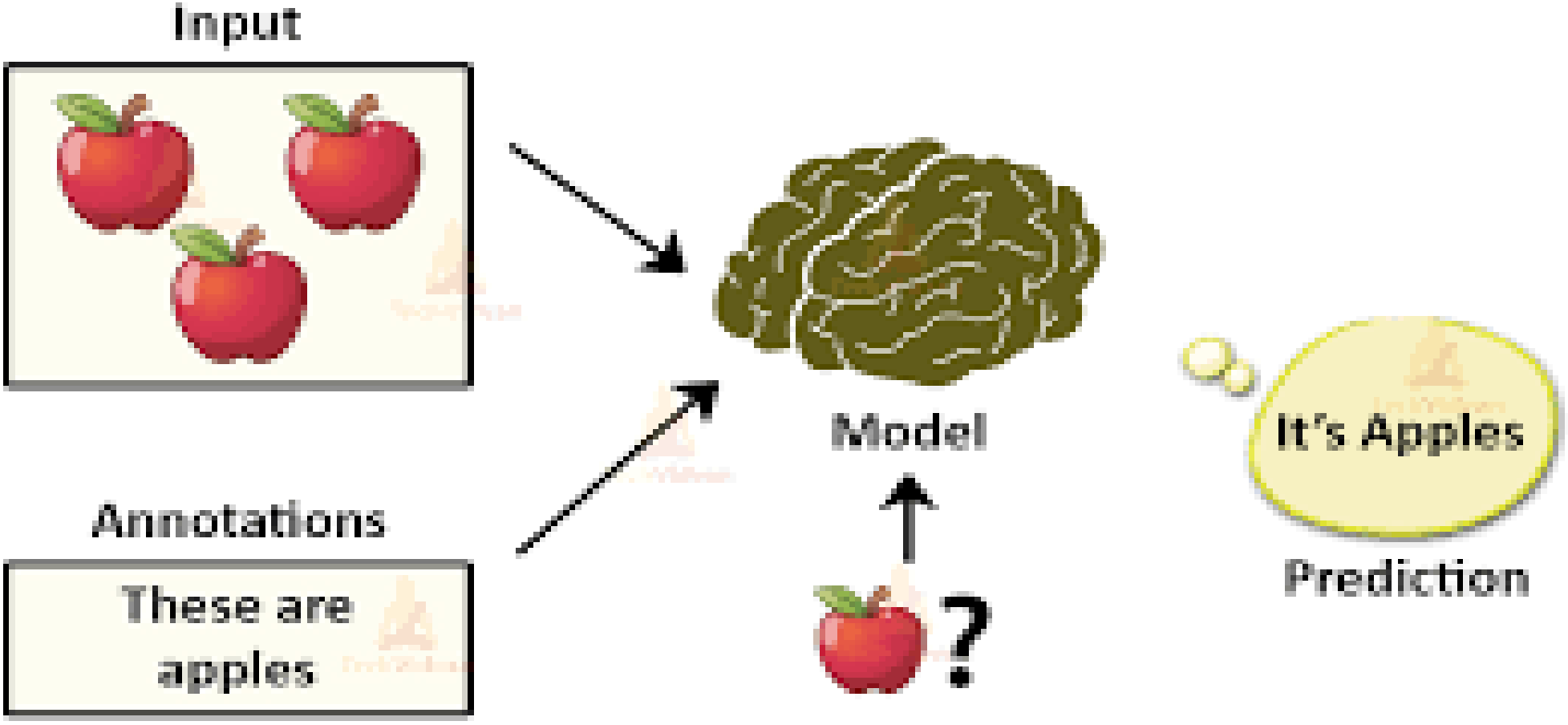
Unsupervised Learning

Reinforcement Learning



simplilearn

Supervised Learning in ML



Tools and Libraries used

- Python
- Numpy
- Scikit-learn
- Matplotlib
- Imblearn
- Pandas

Credit Card Fraud:

- Credit Card fraud means using person’s credit card without his knowledge by means of withdrawing funds or purchase of goods.
- The common ways of fraud:
  - Cloning and skimming.
  - Smishing.
  - Vishing.
  - Indentity theft.

Steps to Develop Credit Card Fraud Classifier in ML:

Our approach to building the classifier is discussed in the steps:

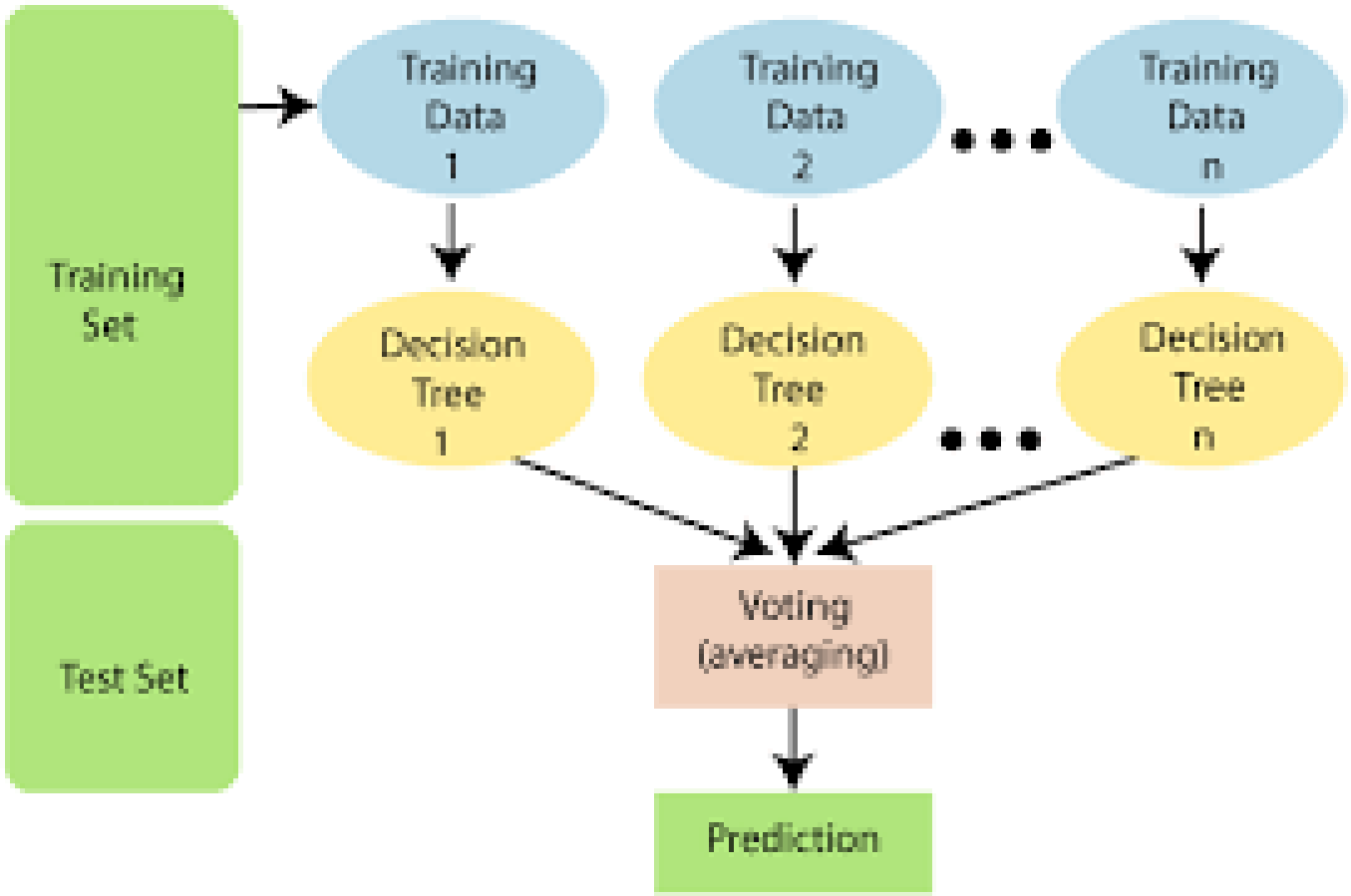
- Perform Exploratory Data Analysis (EDA) on our dataset.
- Apply Machine Learning algorithm to our dataset.
- Train and Evaluate our models on the dataset and find the Accuracy.

Algorithms used:

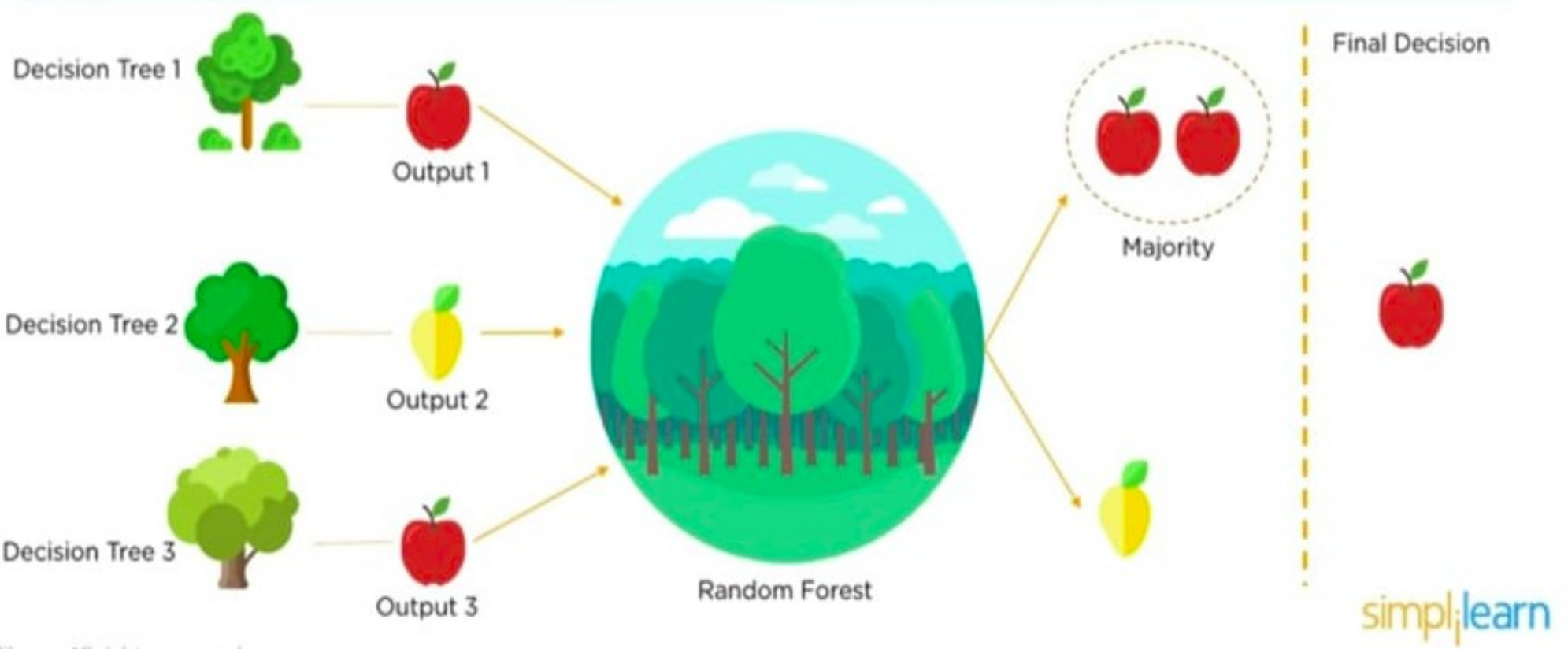
- Random Forest Algorithm:

  A Random Forest is a meta estimator that fits a number of decition tree classifiers on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting.

Random Forest Algarithm:



Random Forest Algarithm:



How does Random Forest Algorithm works:

Step 1:select random k points from the training set.

Step2:Build the decition trees associated with the selected data points.

Step3:Choose the number N for decition trees that you want to build.

Step4:Repeat step 1 & 2.

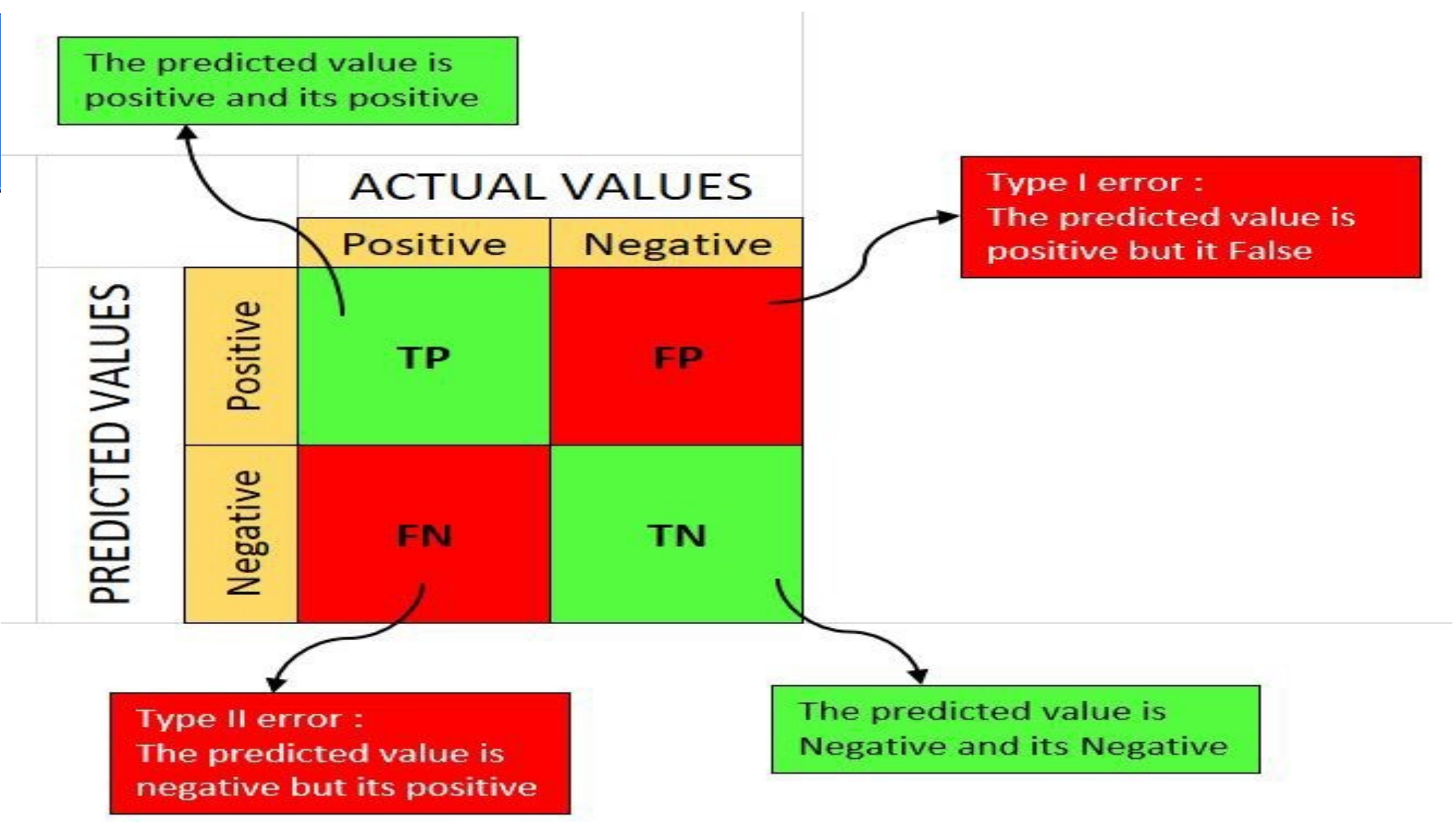
step5:For new data points,find the predictions of each decition tree,and assign the new data points to the category that wins the majority votes.

Why use Random Forest:

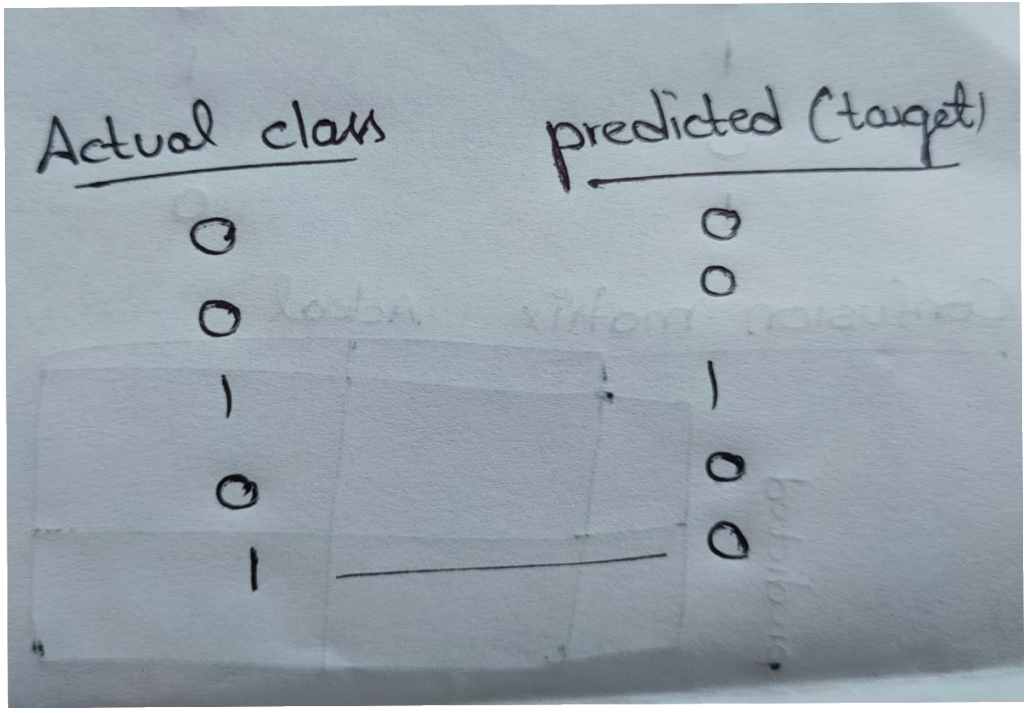
- It takes less training time as compared to other algorithms.
- It predicts output with high accuracy,even for the large dataset it runs efficiently.
- It can also maintain accuracy when a large proportion of data missing.

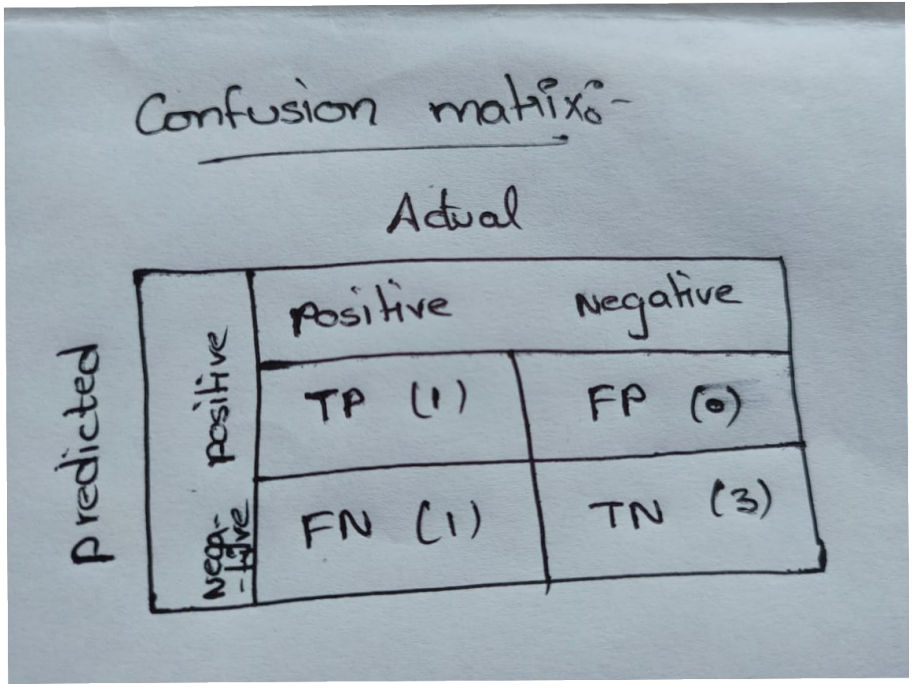
Confusion matrix:

- A confusion matrix is a matrix(table) that can be used to measure the performance of an machine learning algarithm,usually a supervised learning one.
- Each row of the confusion matrix represents the instances of an predicted class and each column represents the instances of a actual class.


The predicted value is positive and its positive
ACTUAL VALUES
Positive
Negative
PREDICTED VALUES
Positive
Negative
TP
FP
FN
TN
Type I error :
The predicted value is positive but it False
Type II error :
The predicted value is negative but its positive
The predicted value is Negative and its Negative

Confusion matrix:





Precision : Precision is used to calculate the model's ability to classify positive values correctly. It answers the question, "When the model predicts a positive value, how often is it right?"

It is the true positives divided by total number of predicted positive values

$$\text{Precision} = \frac{TP}{TP + FP}$$

Accuracy : The accuracy is used to find the portion of correctly classified values.
It tells us how often our classifier is right

It is the sum of all true values divided by total values

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

F1-Score : It is the harmonic mean of Recall and Precision. It is useful when you need to take both Precision and Recall into account

$$\text{F1-Score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

Recall : It is used to calculate the model's ability to predict positive values. "How often does the model actually predict the correct positive values?"

It is the true positives divided by total number of actual positive values

$$\text{Recall} = \frac{TP}{TP + FN}$$

Important steps to Avoid Fraud:

- Never give your card to others.
- Never share your banking details in any online portals and social media.
- Do not write the credit card number or pin number in any places.
- Do not disclose card information and OTP over the phone calls.
- Do not leave the card or receipts unattended.
- Communicate with the card company immediately regarding issues like loss of card or change of address or any doubtful cards.


ANY
QUESTIONS?

Thank you