Introduction to Machine Learning LA Activity Loan Eligibility Prediction using Machine learning

DONE BY:-

TARUN KUMAR ARCOT 1NT18CS174

RAKSHITH R 1NT19CS413

Introduction

This project aims to help the banks to avoid check if the customer who has bank account will repay the loan amount or not.

This project uses Machine learning algorithms to predict the outcome.

Background

- In today's world loan disbursement is one of the major business line for the banks.
- However Loan disbursement can be a risky business if the banks are unable to take an informed decision on whom to give the loan and whom not.
- Banks ask for various information and documents to ascertain the credibility of the client.
- In earlier days, humans used to make the decisions, but the human may use certain emotional judgement even if some of the information is not in the client favour.
- For e.g. even if the client is 55 year old and has only 5 years left got his retirement and if the bank Manager knows him well, the client may be issued the loan.

Machine Learning Steps

1. **Importing packages**
2. **Data Source and quality**
3. **Data Preprocessing**
    1. Data Cleaning
    2. Data Transformation
4. **Machine Leaning Algorithms**
    1. Decision tree Classifier
    2. Naïve Bayes
    3. Random Forest
    4. Logistics Regression
    5. KNN
    6. SVM

NOTE:- The importing packages, data source, quality steps will be shown in the code.

Various libraries used in projects

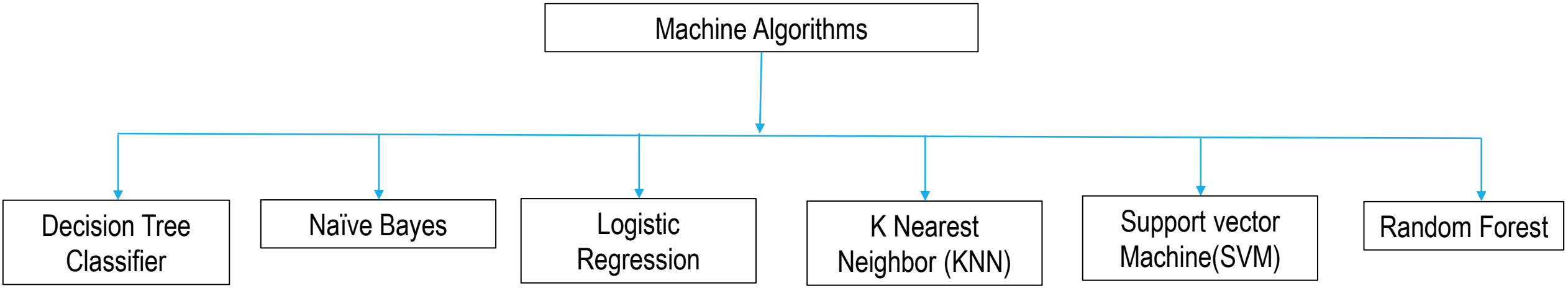
Powerful fast and flexible library for Data Analysis, manipulation and filtering

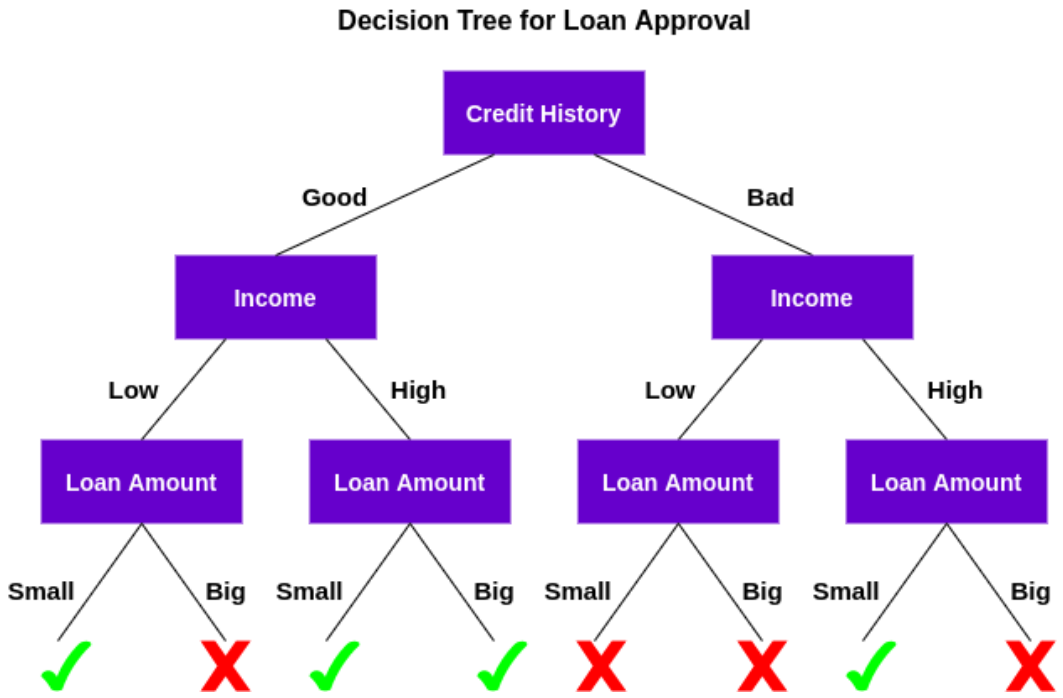
Used to perform mathematical and logical operations on arrays

Comprehensive library for creating static, interactive and animated visualizations

Most useful library with efficient tools for machine learning and statistical modeling including range of supervised and unsupervised learning algorithms

Machine learning Algorithms Used





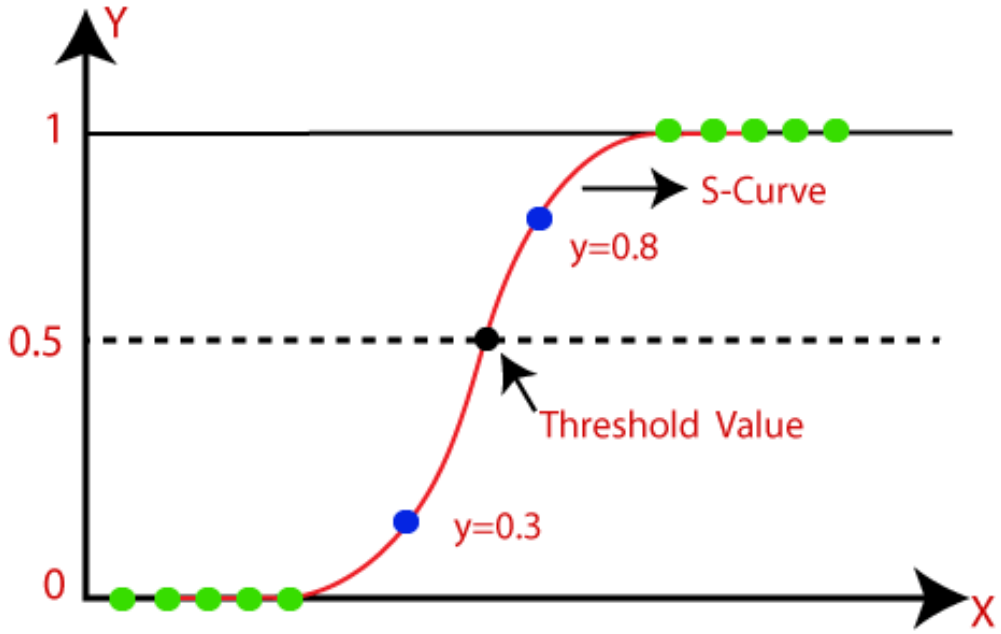
Decision Tree Classifier

- Decision Tree is a **Supervised learning technique** that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems.
- It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules, and each leaf node represents the outcome.
- In a Decision tree, there are two nodes, which are the Decision Node and Leaf Node.

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)}$$

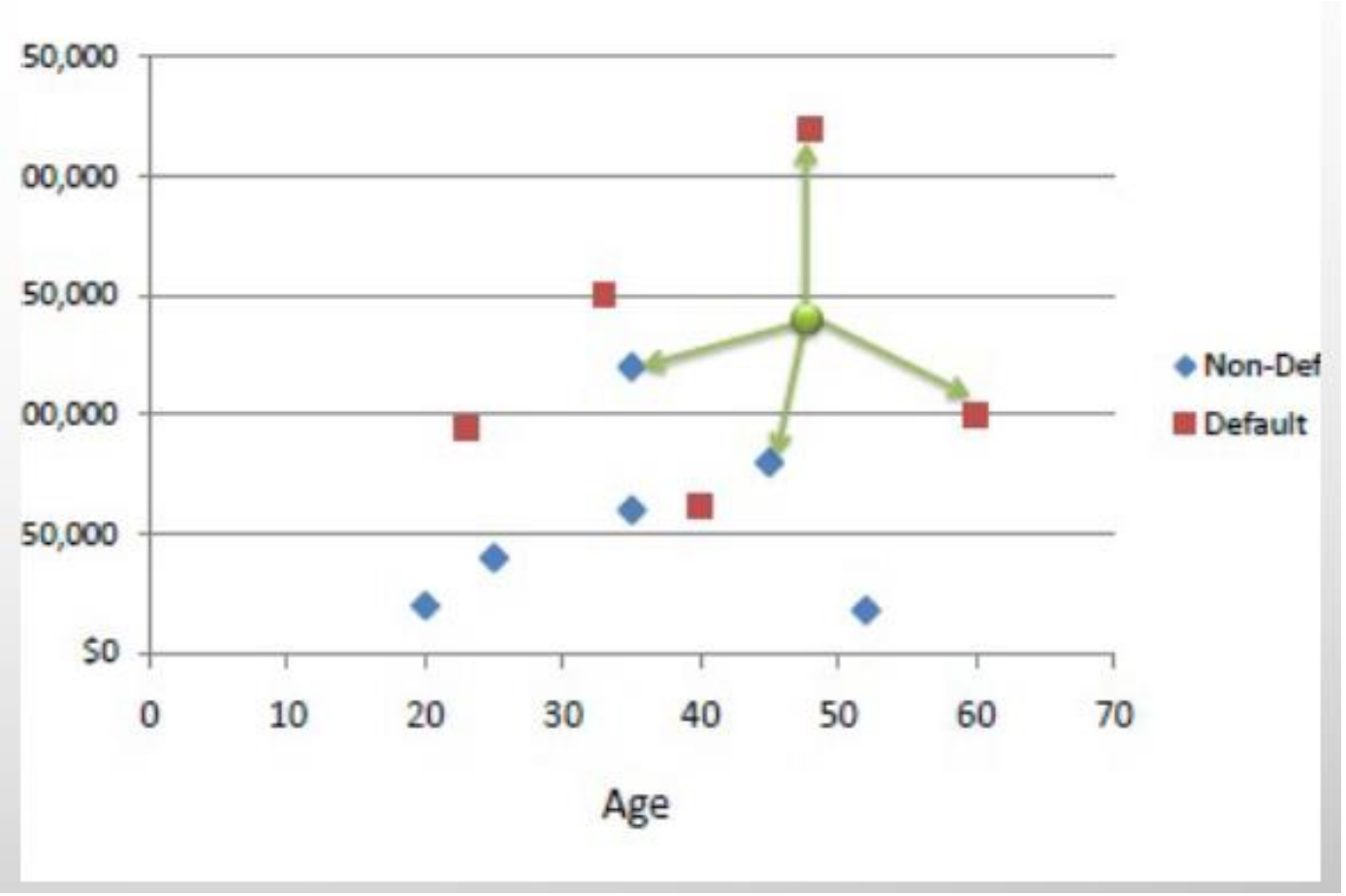
Naïve Bayes

- Naïve Bayes algorithm is a supervised learning algorithm, which is based on **Bayes theorem** and used for solving classification problems.
- It is mainly used in text classification that includes a high-dimensional training dataset.
- The Naïve Bayes algorithm is comprised of two words Naïve and Bayes, Which can be described as:
  - Naïve: It is called Naïve because it assumes that the occurrence of a certain feature is independent of the occurrence of other features. Such as if the fruit is identified on the bases of color, shape, and taste, then red, spherical, and sweet fruit is recognized as an apple. Hence each feature individually contributes to identify that it is an apple without depending on each other.
  - Bayes: It is called Bayes because it depends on the principle of Bayes' Theorem.



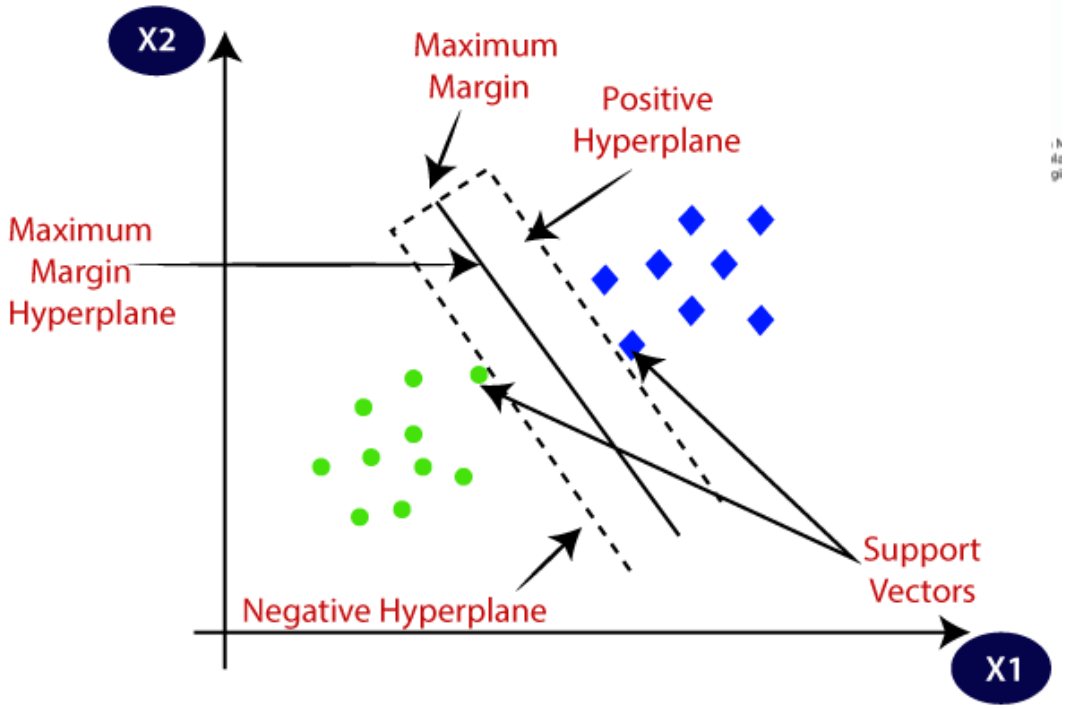
Logistics Regression

- Logistic regression is one of the most popular Machine Learning algorithms, which comes under the Supervised Learning technique.
- It is used for predicting the categorical dependent variable using a given set of independent variables.
- Logistic regression predicts the output of a categorical dependent variable. Therefore, the outcome must be a categorical or discrete value. It can be either Yes or No, 0 or 1, true or False, etc. but instead of giving the exact value as 0 and 1, it gives the probabilistic values which lie between 0 and 1.
- In Logistic regression, instead of fitting a regression line, we fit an "S" shaped logistic function, which predicts two maximum values (0 or 1).



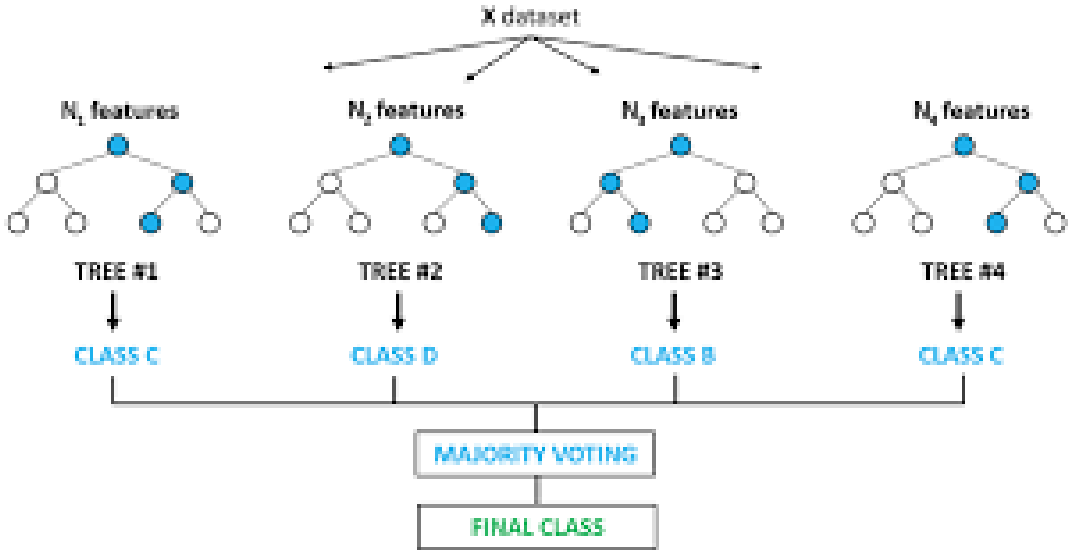
K Nearest Neighbor (KNN)

- K-Nearest Neighbor is one of the simplest Machine Learning algorithms based on Supervised Learning technique.
- K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories.
- K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories.
- K-NN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category by using K- NN algorithm.



Support Vector Machine(SVM)

- Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning.
- The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane.



Random Forest

Random Forest is a supervised learning technique.

It can be used for both Classification and Regression problems in ML.

It is based on the concept of **ensemble learning,** which is a process of *combining multiple classifiers to solve a complex problem and to improve the performance of the model.*

Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset.

Implementation

ALL THE BEST FOR YOUR CODING JOURNEY IN MACHINE LEARNING!!!

Keep coding!!!🤖👍👍

Thank you