

# Classification Algorithms

## Introduction

Classification algorithms are supervised learning techniques used to assign input data into predefined categories or classes. The output of a classification model is discrete, such as yes/no, true/false, or class labels.

## What is Classification?

Classification is used when the target variable represents a category or label. The model learns decision boundaries that separate different classes.

Input Features → Classification Model → Class Label

## Types of Classification Problems

- 1 Binary Classification (two classes)
- 2 Multi-class Classification (more than two classes)
- 3 Multi-label Classification (multiple labels per instance)

## Common Classification Algorithms

- 1 Logistic Regression
- 2 K-Nearest Neighbors (KNN)
- 3 Naive Bayes
- 4 Decision Tree
- 5 Support Vector Machine (SVM)
- 6 Random Forest

### 1. Logistic Regression

Logistic Regression predicts the probability of a class using a sigmoid function. It is widely used for binary classification problems.

Example: Spam vs Not Spam.

Input → Sigmoid Function → Probability → Class

### 2. K-Nearest Neighbors (KNN)

KNN classifies a data point based on the majority class of its nearest neighbors.

Example: Handwritten digit recognition.

New Point → Find K Nearest → Majority Vote → Class

### 3. Naive Bayes

Naive Bayes is a probabilistic classifier based on Bayes' Theorem. It assumes independence among features.

Example: Text classification.

### 4. Decision Tree

Decision Trees split data into branches based on feature conditions, forming a tree-like structure.

Example: Loan approval system.

Feature Test → Branch → Decision

## 5. Support Vector Machine (SVM)

SVM finds an optimal hyperplane that separates classes with maximum margin.

Example: Face recognition.

## 6. Random Forest

Random Forest is an ensemble method that combines multiple decision trees to improve accuracy and reduce overfitting.

Example: Disease prediction.

## Classification Evaluation Metrics

- 1 Accuracy
- 2 Precision
- 3 Recall
- 4 F1-score
- 5 Confusion Matrix

Predicted		
	Positive	Negative
Actual	TP	FN
	FP	TN

## Advantages

- 1 Handles categorical output effectively
- 2 Widely used in real-world applications
- 3 Interpretable models available

## Disadvantages

- 1 Sensitive to noisy data
- 2 Class imbalance issues
- 3 Some models are computationally expensive

## Real-Life Example

In medical diagnosis, classification algorithms are used to predict whether a patient has a particular disease based on symptoms and test results.

## Summary

Classification algorithms are essential supervised learning techniques for assigning class labels. Selecting the appropriate algorithm depends on data size, complexity, and accuracy requirements.