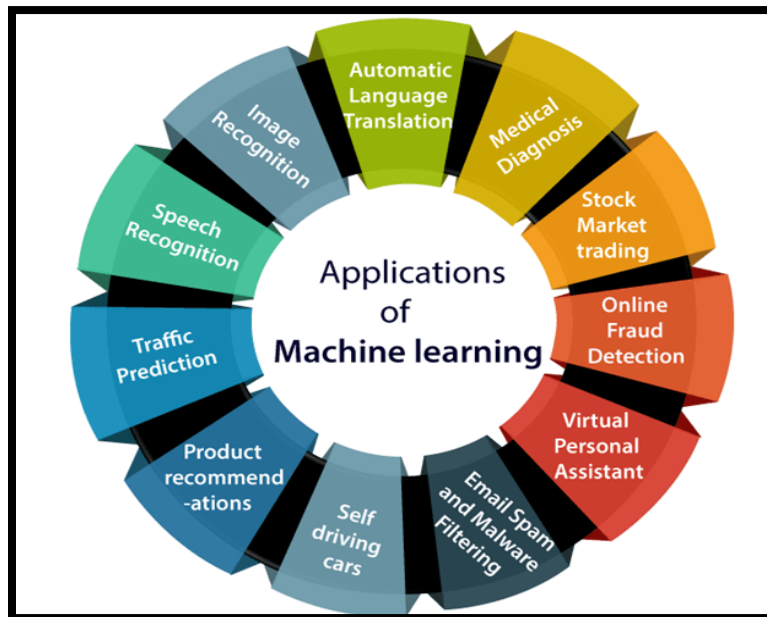


SUPERVISED LEARNING

**Explaining supervised learning algorithms, including
Regression and Classification**

INTRODUCTION

AI driven innovations have completely transformed our lives by impacting almost every industry. Technological advancements in healthcare, businesses, education, entertainment industries now extensively depend on AI. From personalised recommendations to disease detection to self-driving cars , Machine Learning algorithms are at the heart of these innovations. In machine learning, supervised learning is one such paradigm that takes insights from labelled datasets and produces predictions.



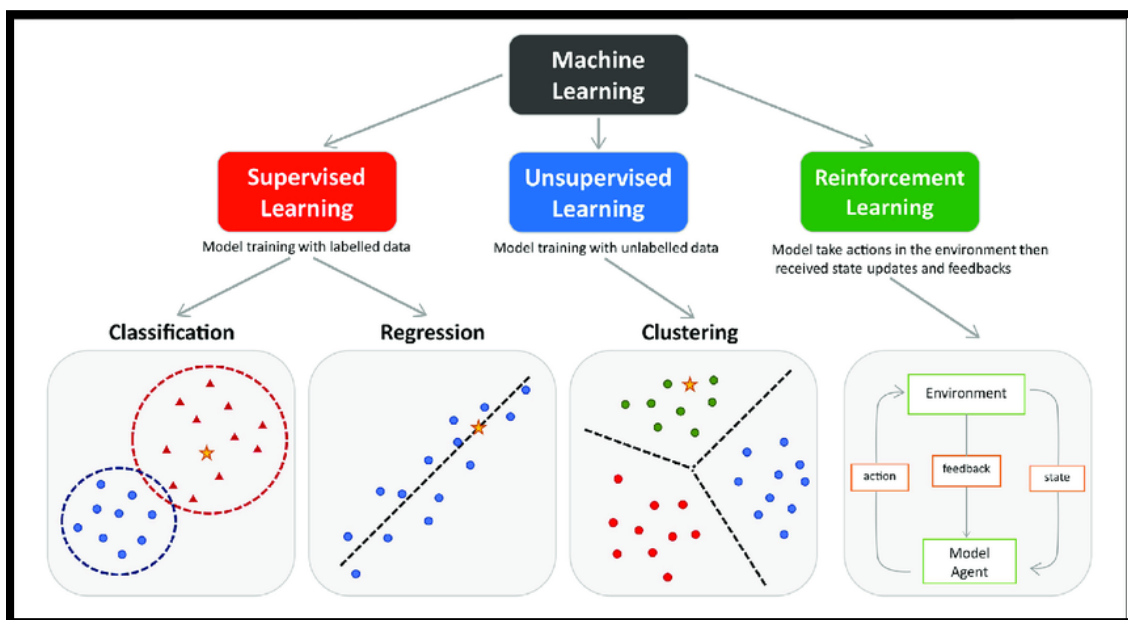
In this article, we will have a detailed analysis on Supervised Learning, exploring its major types and popularly used models.

WHAT IS MACHINE LEARNING?

Machine learning is a subset of Artificial Intelligence . It is the study of statistical and computational techniques which enables a computer to learn from data. The system learns from the data, identifies patterns and generates predictions.

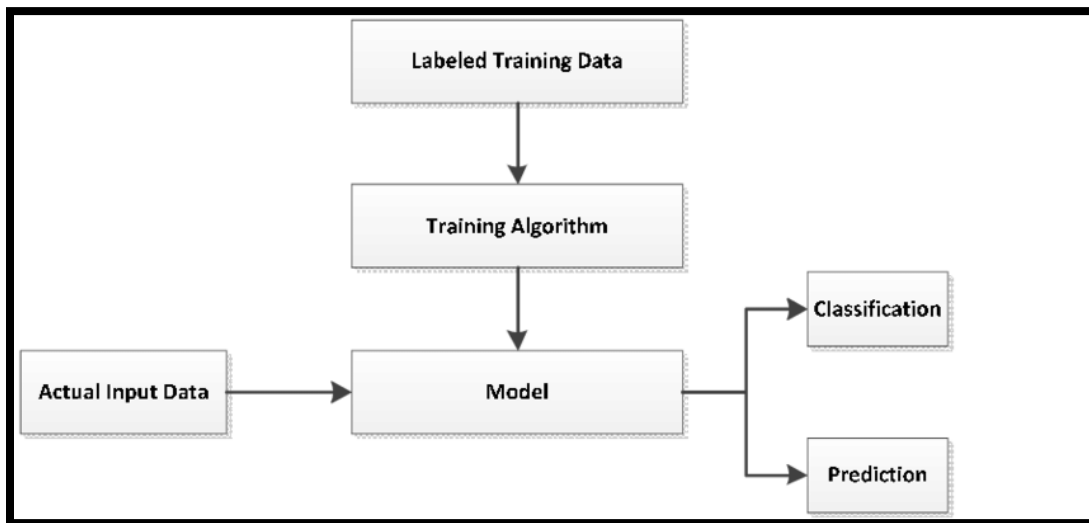
MACHINE LEARNING LEARNING PARADIGMS

Machine Learning is categorised into three broad categories based on the type of data input and learning ability. These categories are then further divided into different classes.



SUPERVISED LEARNING PARADIGM

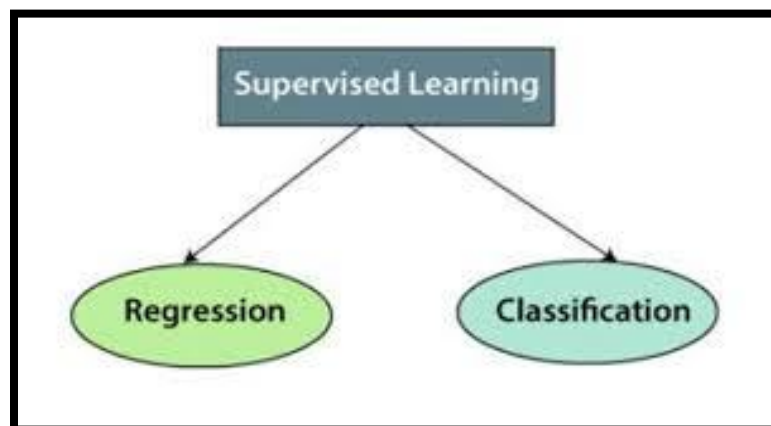
In this type of learning the algorithms are trained with labelled dataset. The labelled dataset contains input variables (features) and their associated outputs (labels). The model learns the relationship between the features and labels and then uses the learned techniques for making predictions .



Block Diagram of Supervised Learning

TYPES OF SUPERVISED LEARNING ALGORITHMS

The two types are as follows:



- **Classification:**

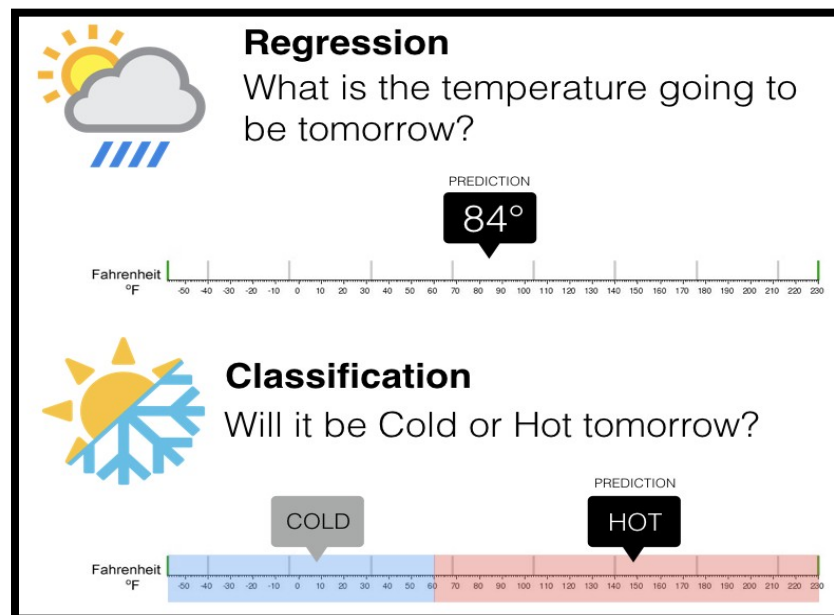
The classification model divides the data into categories and then predicts the probability of a certain type of data falling into the predefined categories. Some examples of this model include :

- filtration of email into 'spam' and 'not spam'.
- Identifying gender as 'Male' or 'Female'
- Identifying a plants specie
- Based on temperature data predict it will 'rain' or 'not rain'

- **Regression:**

The regression model predicts numerical values against the provided input parameters. Some examples of this model include:

- Predicting house prices
- Predicting students marks
- Predicting daily temperature
- Predicting sales



Example of Regression and Classification

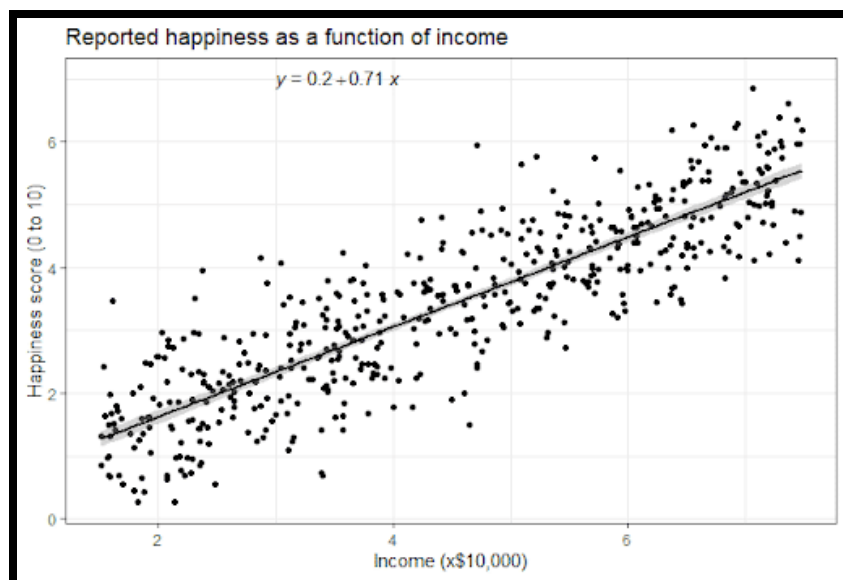
POPULAR REGRESSION AND CLASSIFICATION MODELS

1. Linear Regression

Linear regression is one of the most widely used supervised learning algorithms. which predicts a continuous numeric value . The model is fed with a training dataset and then it is tested upon some unseen data to see how well it has learned to make predictions.

Working:

The model splits the data into 'x' and 'y' variables with 'x' being the independent variable and 'y' being the target or the dependent variable. The target variable can be dependent on one or more variables. The relationship between both types of variables is approximately linear. Mathematically, the linear regression model works on the equation of line $y=mx+c$. It tries to fit a line through the data points to make predictions. The line that fits the best on the data points is known as the **regression line**.



Applications Of Linear Regression:

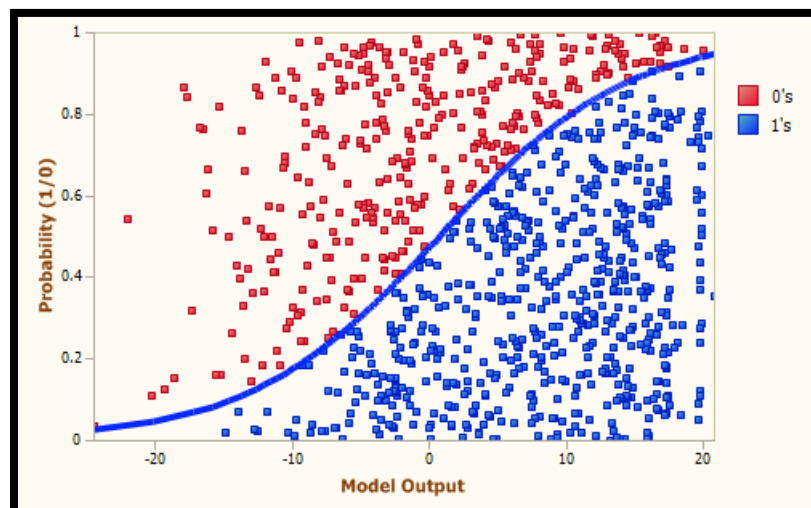
- In marketing, it is used for predicting the impact of advertising expenditure on a product's sales
- In real estate agency, it can be employed in giving house price predictions based on factors like number of room, area , location etc
- In supply chain management, it can be used to estimate the demand for a certain product based on historical sales data.

2. Logistic Regression

The logistic regression model outputs results in a discrete format. It predicts the probability of a data instance belonging to a particular category. It is mostly used in binary classification problems.

Working:

The model analyses the target variable and makes some classes . The target variable is in the form of predefined categories. Each set of independent variables predicts a class. The classes are in the form of 0 or 1 where both represent two different categories.

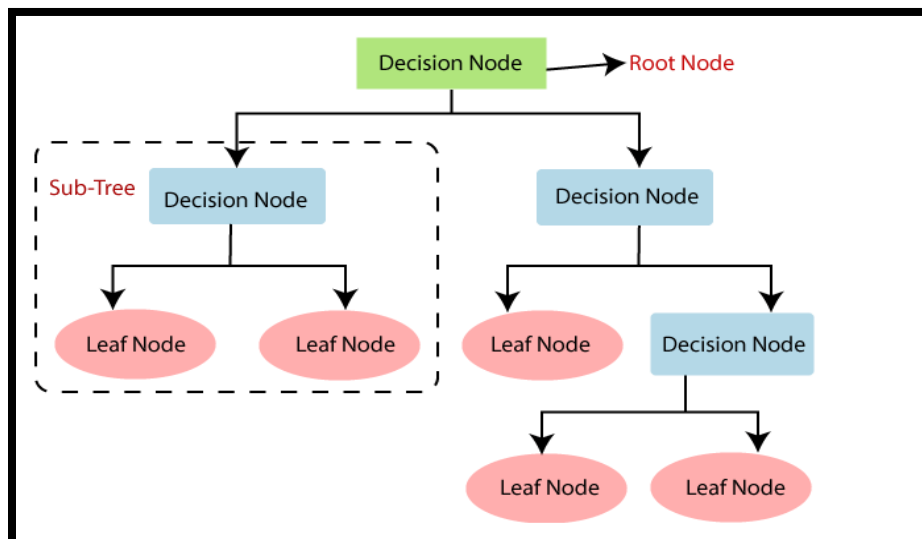


Applications Of Logistic Regression:

- In the healthcare industry, for predicting whether a person suffers from diabetes or not based on factors like age, cholesterol levels, blood pressure, glucose levels etc
- In sports, predicting a team's success or failure. For example, in cricket factors for prediction can be batting score, fielding stats, bowling average etc
- In education, it can be used to analyse whether a student will pass or fail an exam.

3. Decision Trees:

Decision tree is another popular supervised learning algorithm that is used for both classification and regression problems. It is a tree-like data structure which works by recursively splitting the dataset by evaluating the features of the dataset. It is a graphical representation for getting all the possible solutions to a problem.



Working:

The internal nodes of the tree represent the features . These are the decision nodes. On each decision node the tree is further expanded into leaf nodes which represent a particular outcome. The branches of the tree represent the conditions on which a subtree is splitted. A decision evaluates a condition (yes/no) and based on the answer it further splits the tree into subtrees.

For Classification: Decision trees when used for classification problems contain leaf nodes which represent the final decision that is a particular class of the target variable. The decision rules are simple and correspond to all the unique values of features chosen for reaching the final decision.

For Regression: In regression problems , decision trees contain leaf nodes which represent the final decision that is a continuous numerical value of the target variable. The decision rules are simple and correspond to all the unique values of features chosen for reaching the final decision.

Applications Of Decision Trees:

- In fraud detection decision trees are popularly used because of the models ability to handle imbalance data.
- In all of the applications where linear or logistic regression models are applied , decision trees can also be used as they have the tendency to handle both regression and classification tasks.

4. KNN(K - Nearest Neighbours):

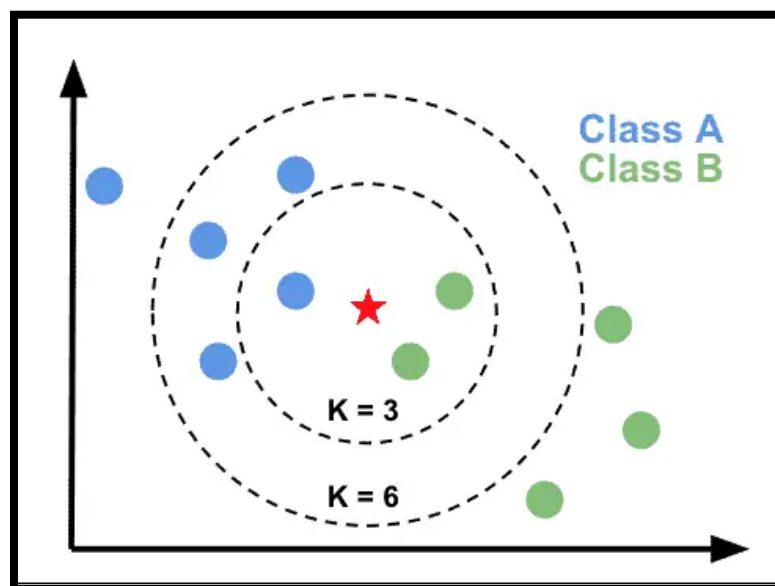
KNN is one of the simplest ML algorithms which can also be used for both classification and regression , but it is mostly used for classification tasks. The KNN algorithm makes predictions by looking for the neighbouring data points (K-Nearest Neighbours) of a particular point and then chooses the majority class. It associates that particular data point to the class of nearest data points.

Working:

The algorithm takes an input parameter 'K' which is the standard of choosing a fixed number of neighbours . This parameter is entered by the user. Then for calculating the nearest proximity , distance metric such as Euclidean Distance is used.

For Classification: KNN works by determining the majority class among the nearest neighbours for a particular data point.

For Regression: For making predictions, KNN determines the average of the nearest neighbours for a particular data point.



Applications Of KNN:

- In image recognition , KNN is used for detecting an object as it can identify patterns in data by comparing it with nearest data points.
- In recommendation systems, KNN is suitable as it gives recommendations based on the preferences of other similar users.

CONCLUSION

In the field of Machine learning , the supervised learning paradigm is evolving rapidly because of its dynamic applications. There are a variety of algorithms offered by this branch of ML and each of them has its own technical capability be it linear or logistic regression, decision trees or KNN or others. But the most important job is the selection of a suitable model.