

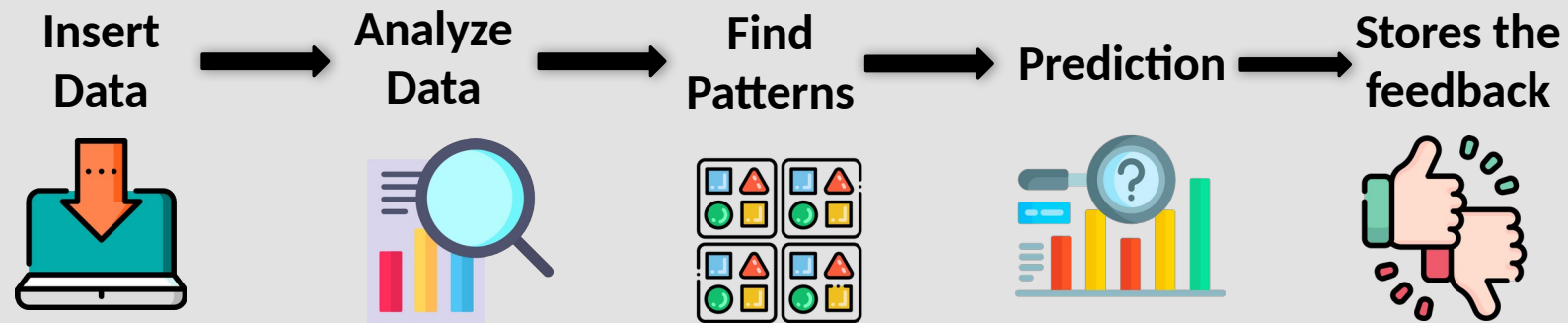


DAY 9

Instructor: Balu Mohandas Menon

Christian B. Wiberg
Philip Jess Teining

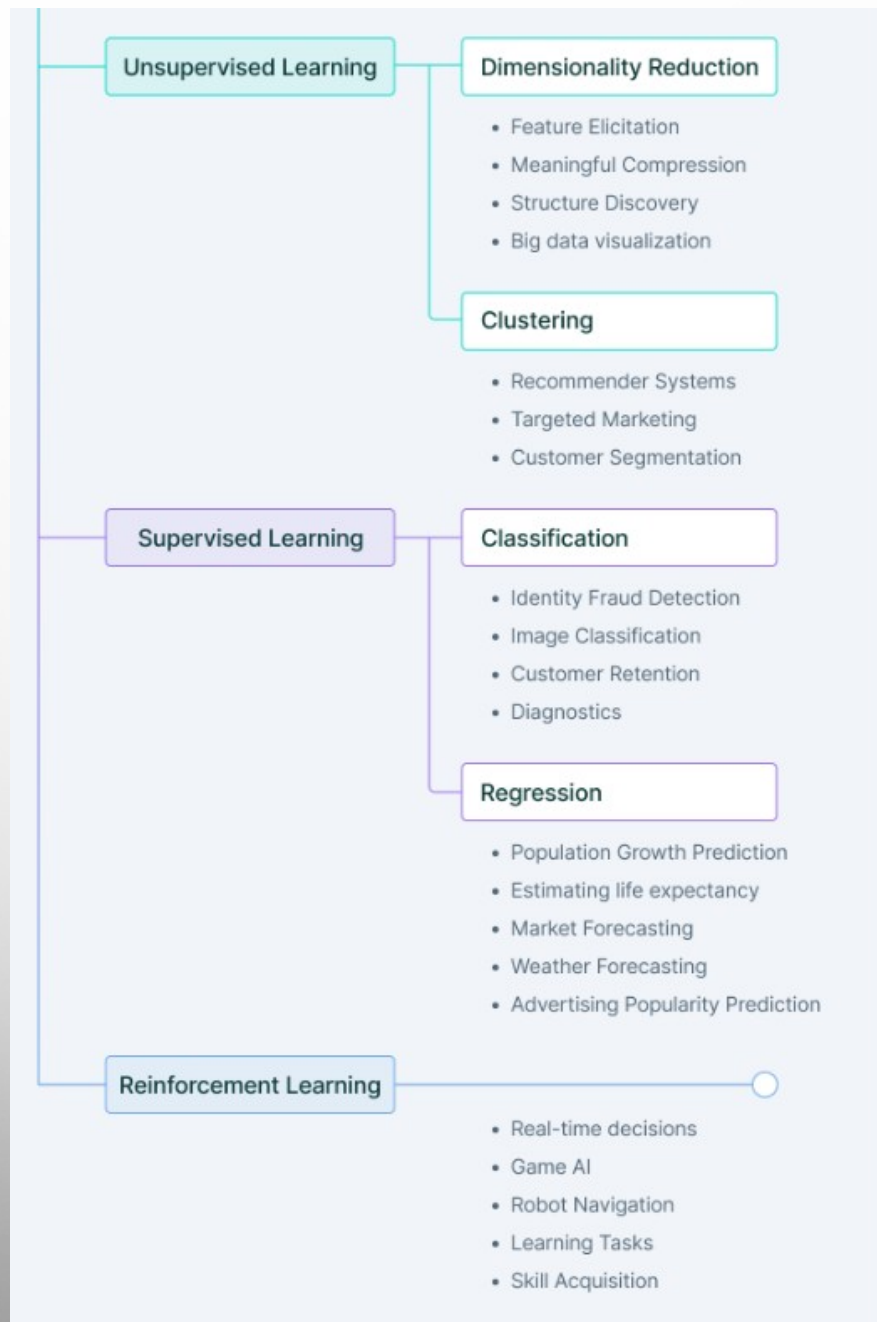
HOW DOES ML WORK?



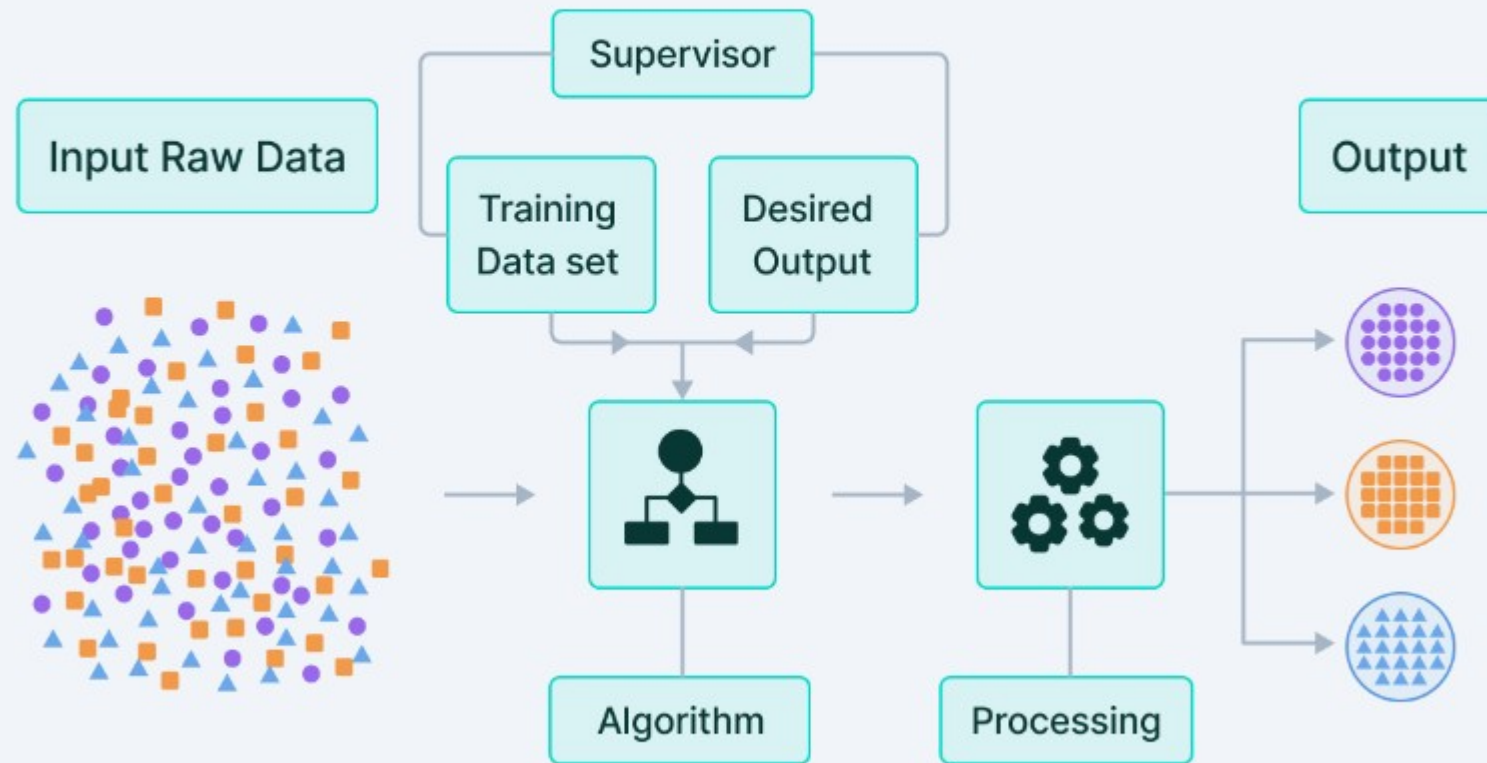
<https://data-flair.training/blogs/machine-learning-tutorial/>

TYPES OF ML

- Supervised learning
- Unsupervised learning
- Reinforcement learning



Supervised Learning



V7 Labs

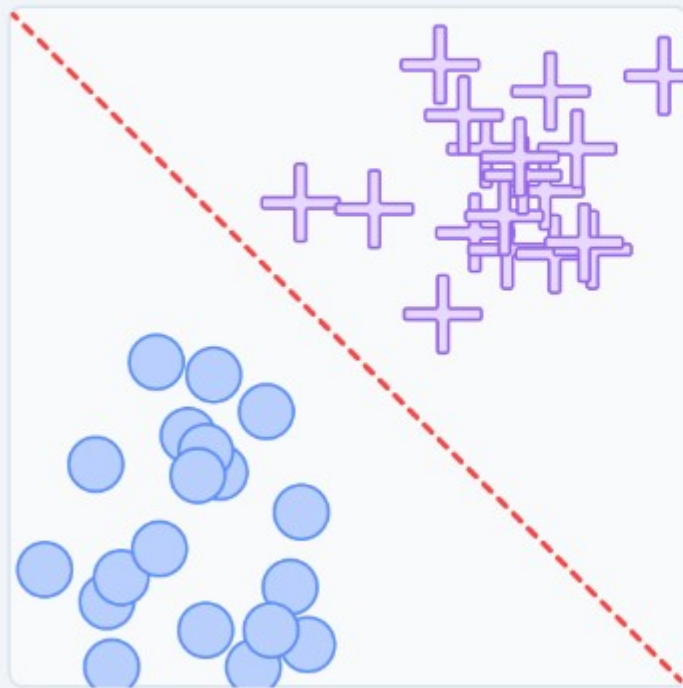
SUPERVISED MACHINE LEARNING METHODS

- **Classification** : Classification refers to taking an input value and mapping it to a discrete value.
 - Output typically consists of categories, ex: whether it is going to rain today or not.
 - **Regression** : predicted output values are real numbers.
 - Predicting the price of a house or the trend in the stock price at a given time, etc.

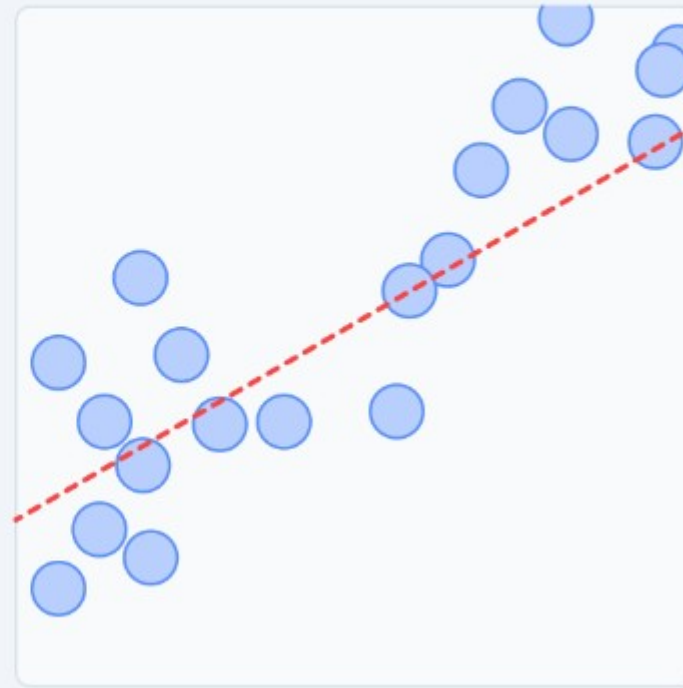
UNSUPERVISED MACHINE LEARNING METHODS

- **Clustering** : Clustering is the type of Unsupervised Learning where we find hidden patterns in the data based on their similarities or difference.
- **Association** : Association is the kind of Unsupervised Learning where we can find the relationship of one data item to another data item.
- e.g., understanding consumers' habits regarding our products can help us develop better cross-selling strategies.

Classification

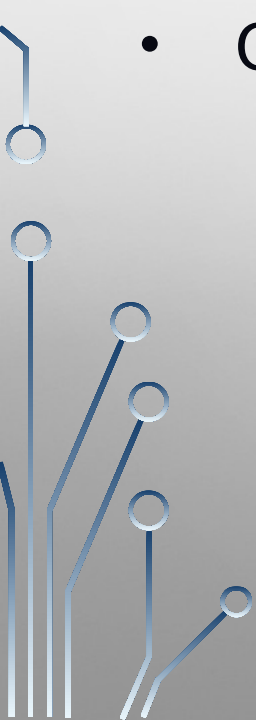


Regression



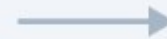


SUPERVISED MACHINE LEARNING APPLICATIONS

- Predictive analytics (house prices, stock exchange prices, etc.)
 - Text recognition
 - Spam detection
 - Customer sentiment analysis
 - Object detection (e.g. face detection)
- 



Supervised Learning



Labeled Data

Unsupervised Learning


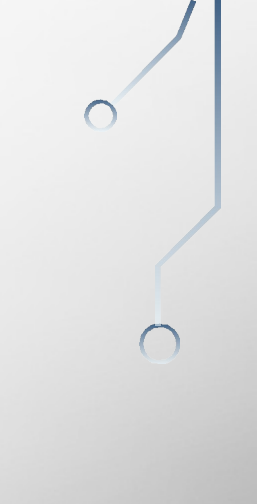
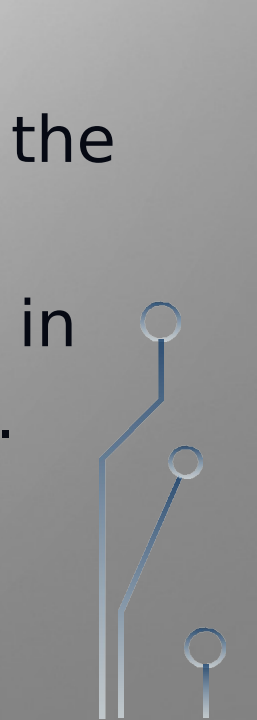


Unlabeled Data





DATA

- ***Supervised Learning*** learns from the training dataset by iteratively making predictions on the data and adjusting for the correct answer.
 - Supervised techniques deal with labeled data where the output data patterns are known to the system.
 - ***Unsupervised Learning*** models work on their own to discover the inherent structure of unlabeled data.
 - The unsupervised learning algorithm works with unlabeled data, in which the output is based solely on the collection of perceptions.
- 
- 
- 

Supervised learning

Input data is labeled

Has a feedback mechanism

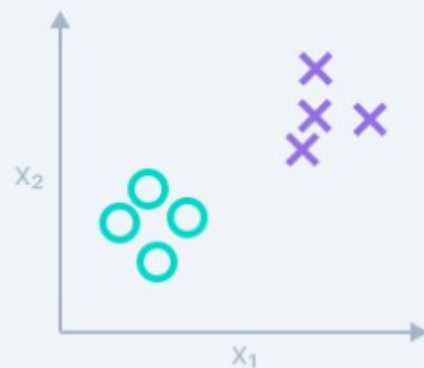
Data is classified based on the training dataset

Divided into Regression & Classification

Used for prediction

Algorithms include: decision trees, logistic regressions, support vector machine

A known number of classes



Unsupervised learning

Input data is unlabeled

Has no feedback mechanism

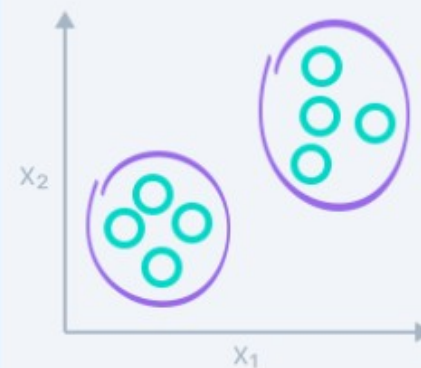
Assigns properties of given data to classify it

Divided into Clustering & Association

Used for analysis

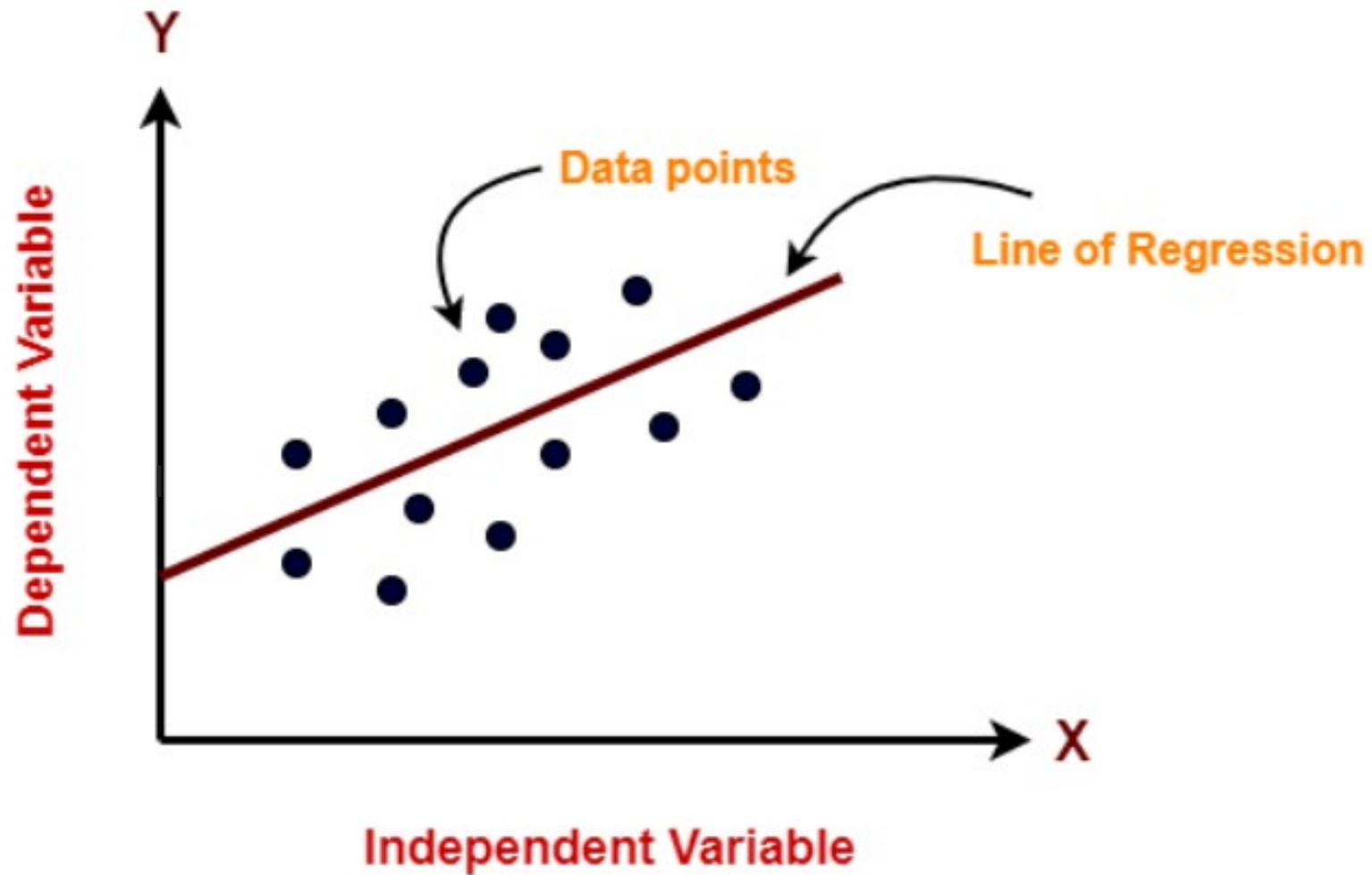
Algorithms include: k-means clustering, hierarchical clustering, apriori algorithm

A unknown number of classes

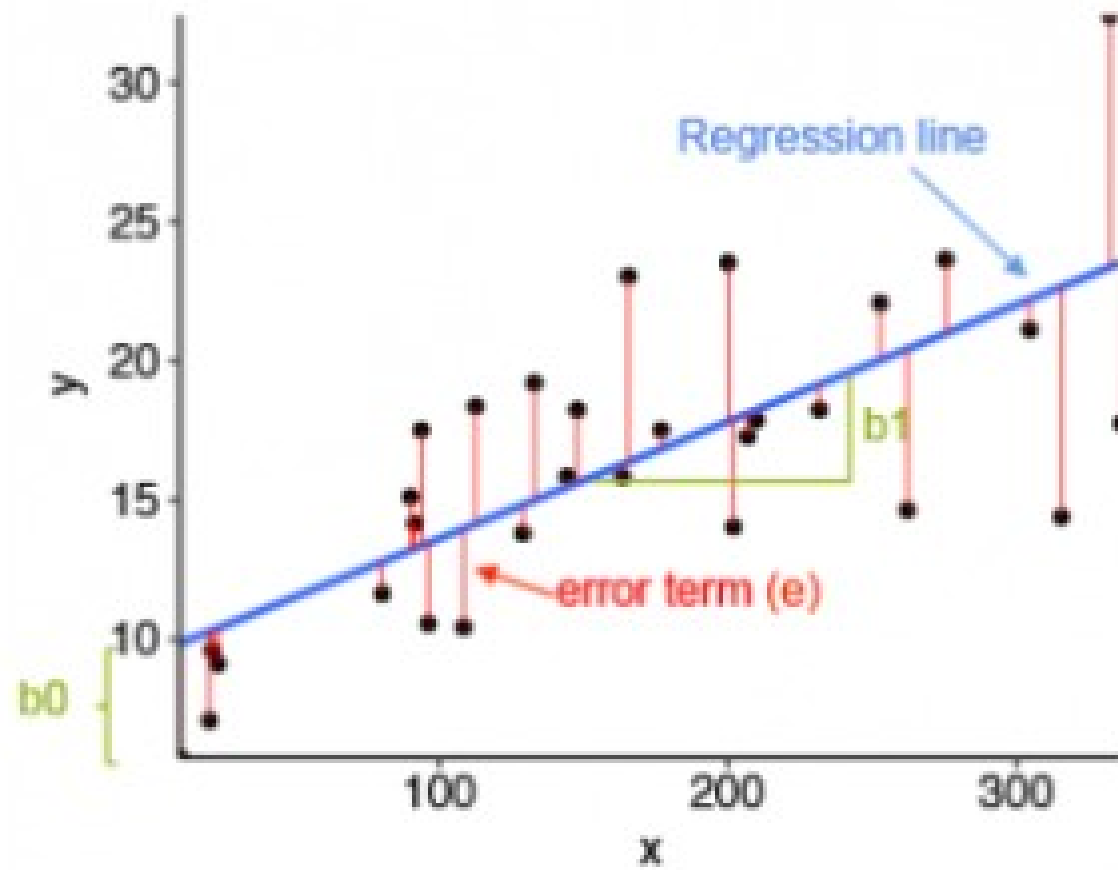


REGRESSION TECHNIQUES

- Regression is a technique for investigating the relationship between independent variables or features and a dependent variable or outcome. It's used as a method for predictive modelling in **machine learning**, in which an algorithm is used to predict continuous outcomes. Regression algorithms include:
 - 1. Linear Regression
 - 2. Decision Tree
 - 3. Support Vector Regression
 - 4. Lasso Regression
 - 5. Random Forest



In the figure above, on X-axis is the independent variable and on Y-axis is the output. The regression line is the best fit line for a model. And our main objective in this algorithm is to find this best fit line.



DECISION TREE

- The decision tree models can be applied to all those data which contains numerical features and categorical features.
- Decision trees are good at capturing non-linear interaction between the features and the target variable.
- Decision trees somewhat match human-level thinking so it's very intuitive to understand the data.

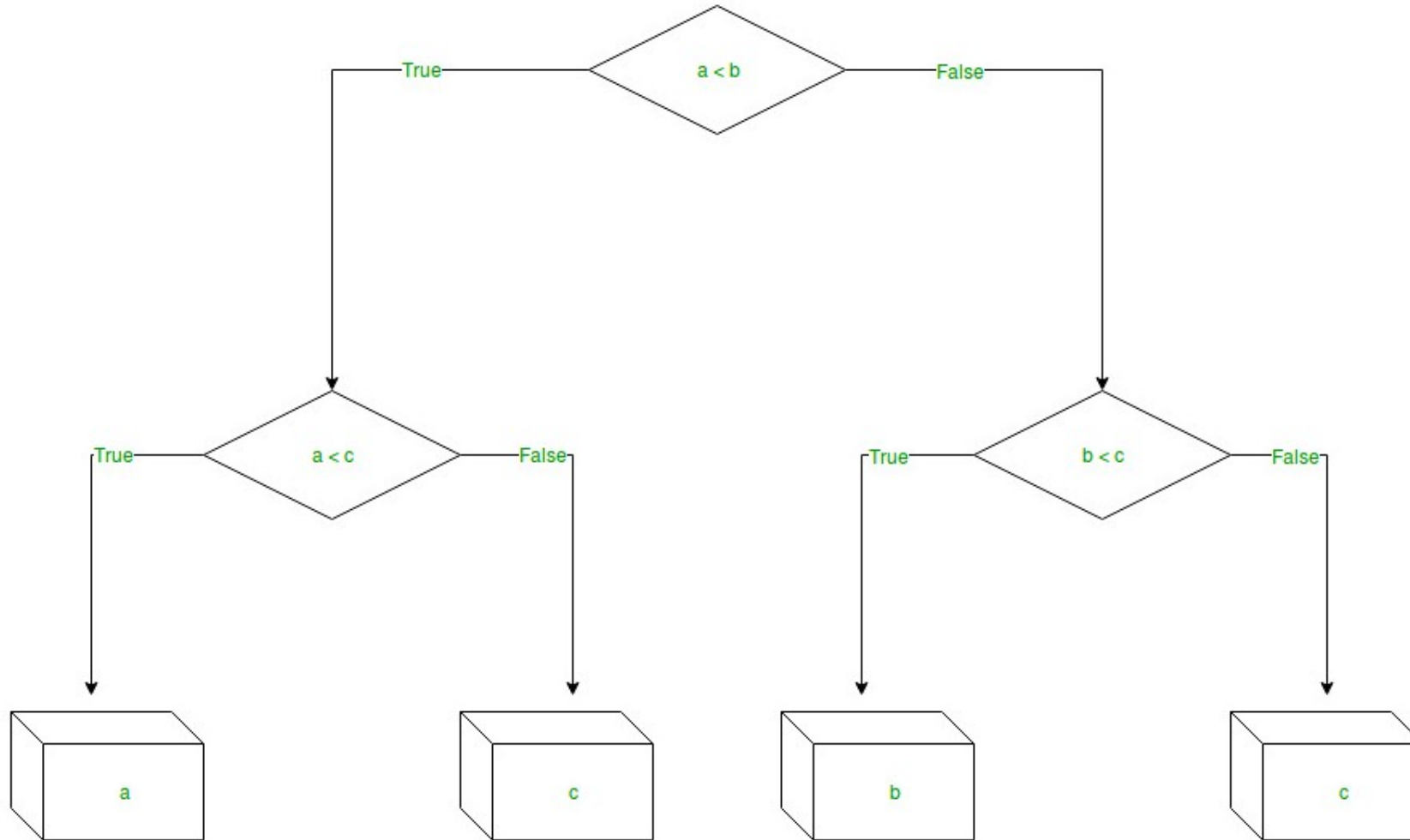
DECISION TREE

- **Decision Tree** is a decision-making tool that uses a flowchart-like tree structure or is a model of decisions and all of their possible results, including outcomes, input costs, and utility.
Decision-tree algorithm falls under the category of supervised learning algorithms. It works for both continuous as well as categorical output variables.

DECISION TREE

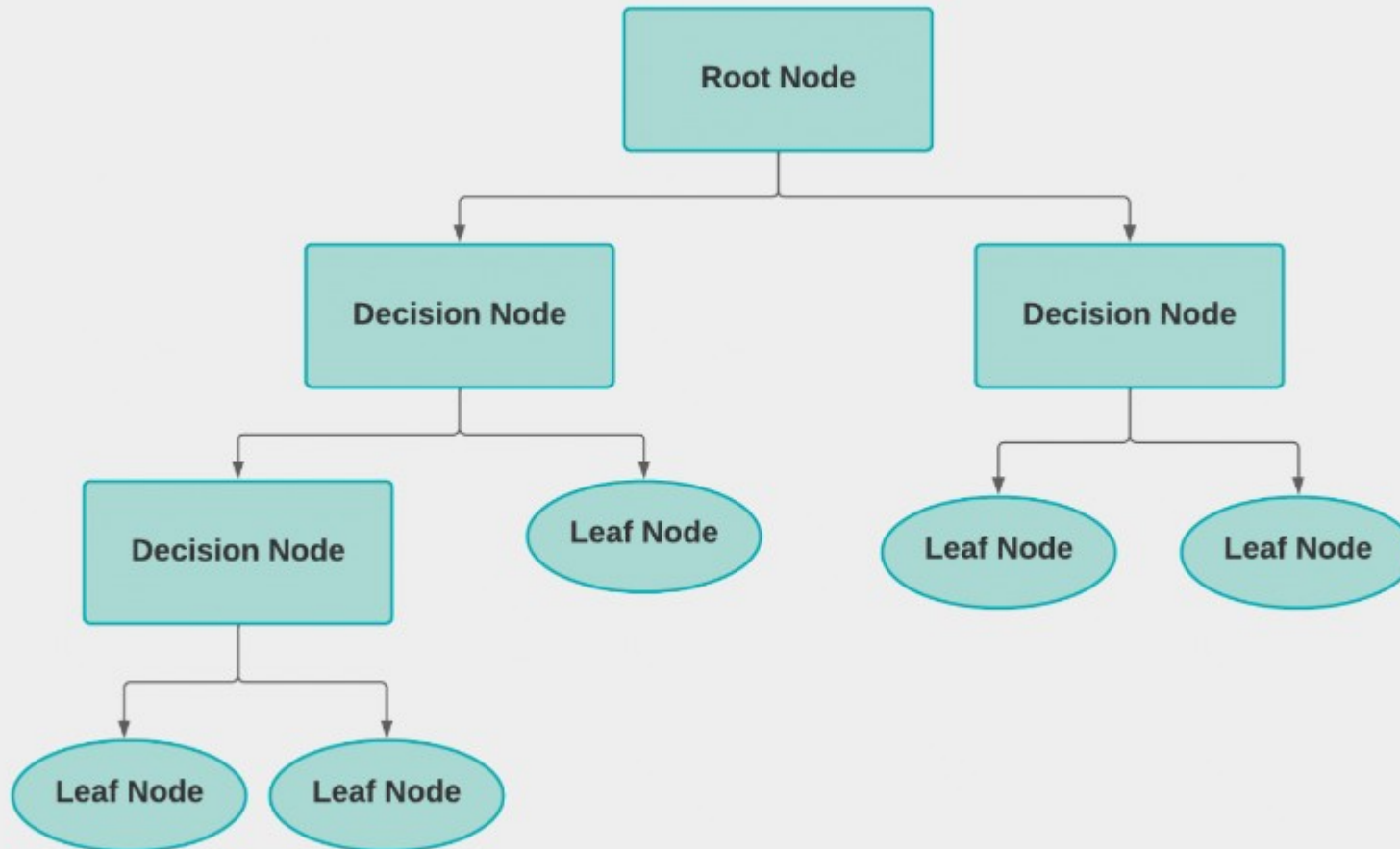
- The branches/edges represent the result of the node and the nodes have either:
 1. Conditions [Decision Nodes]
 2. Result [End Nodes]

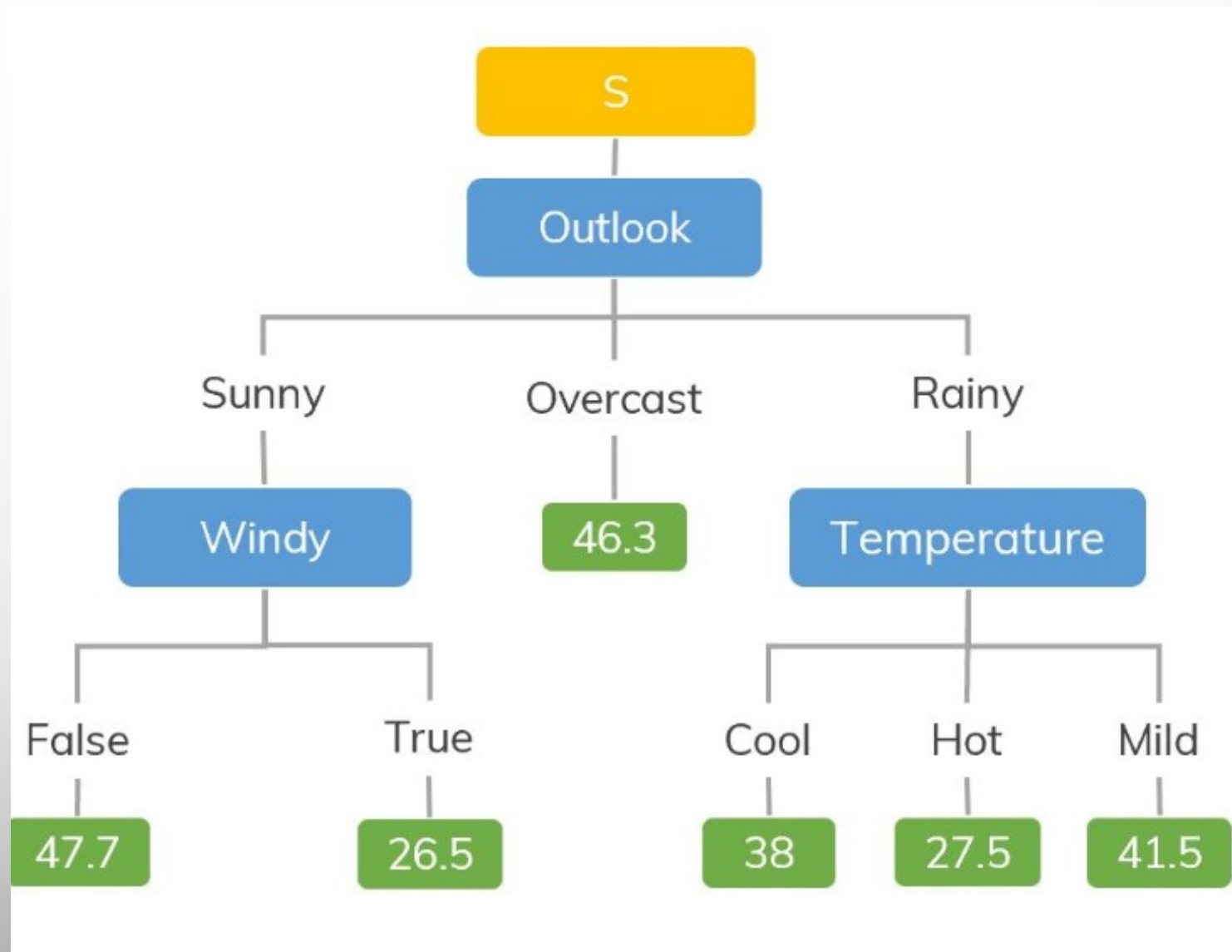
The branches/edges represent the truth/falsity of the statement and take makes a decision based on that in the example below which shows a decision tree that evaluates the smallest of three numbers:



Decision Trees

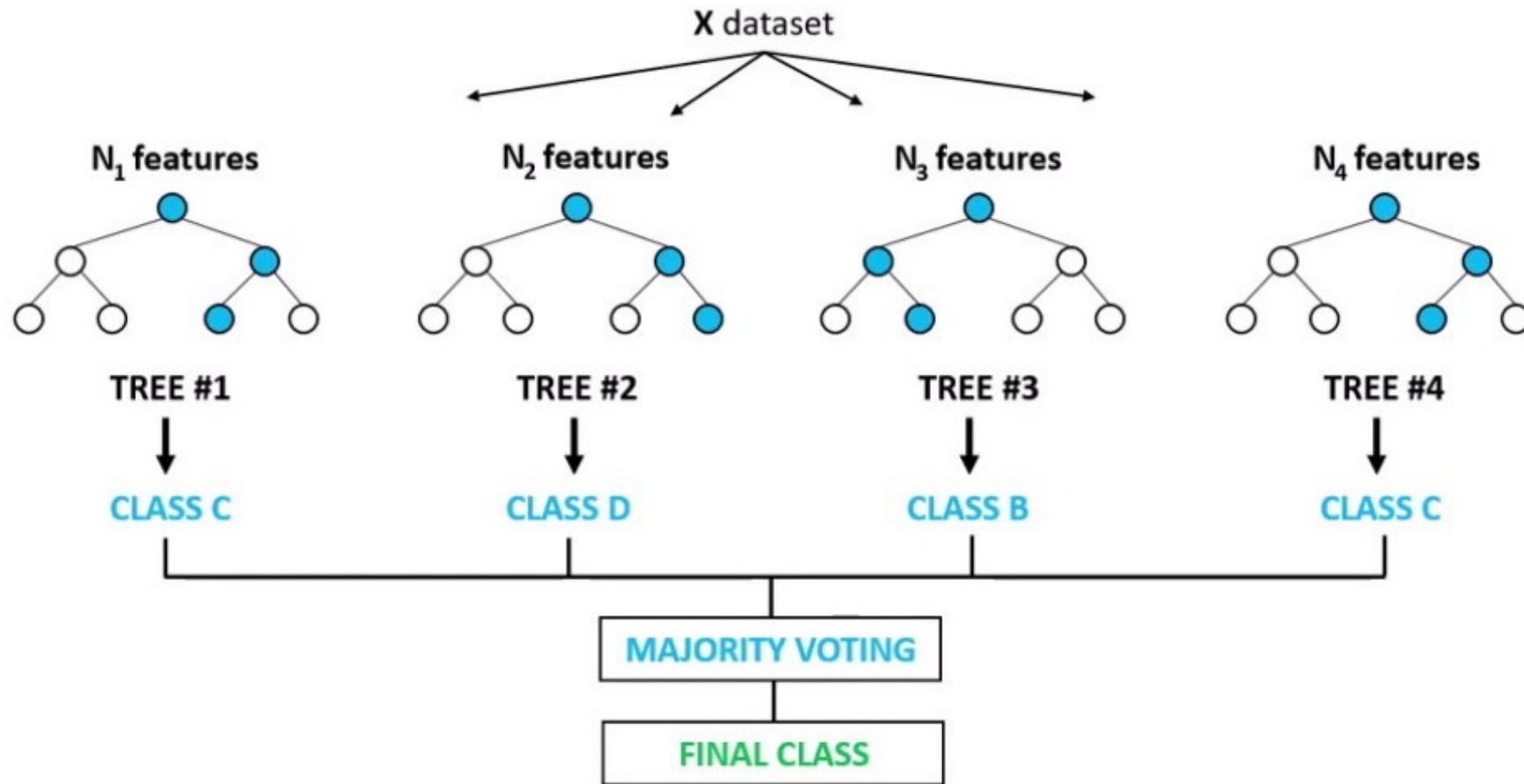
datagy.io





For example, if we are classifying how many hours a kid plays in particular weather then the decision tree looks like somewhat this

Random Forest Classifier



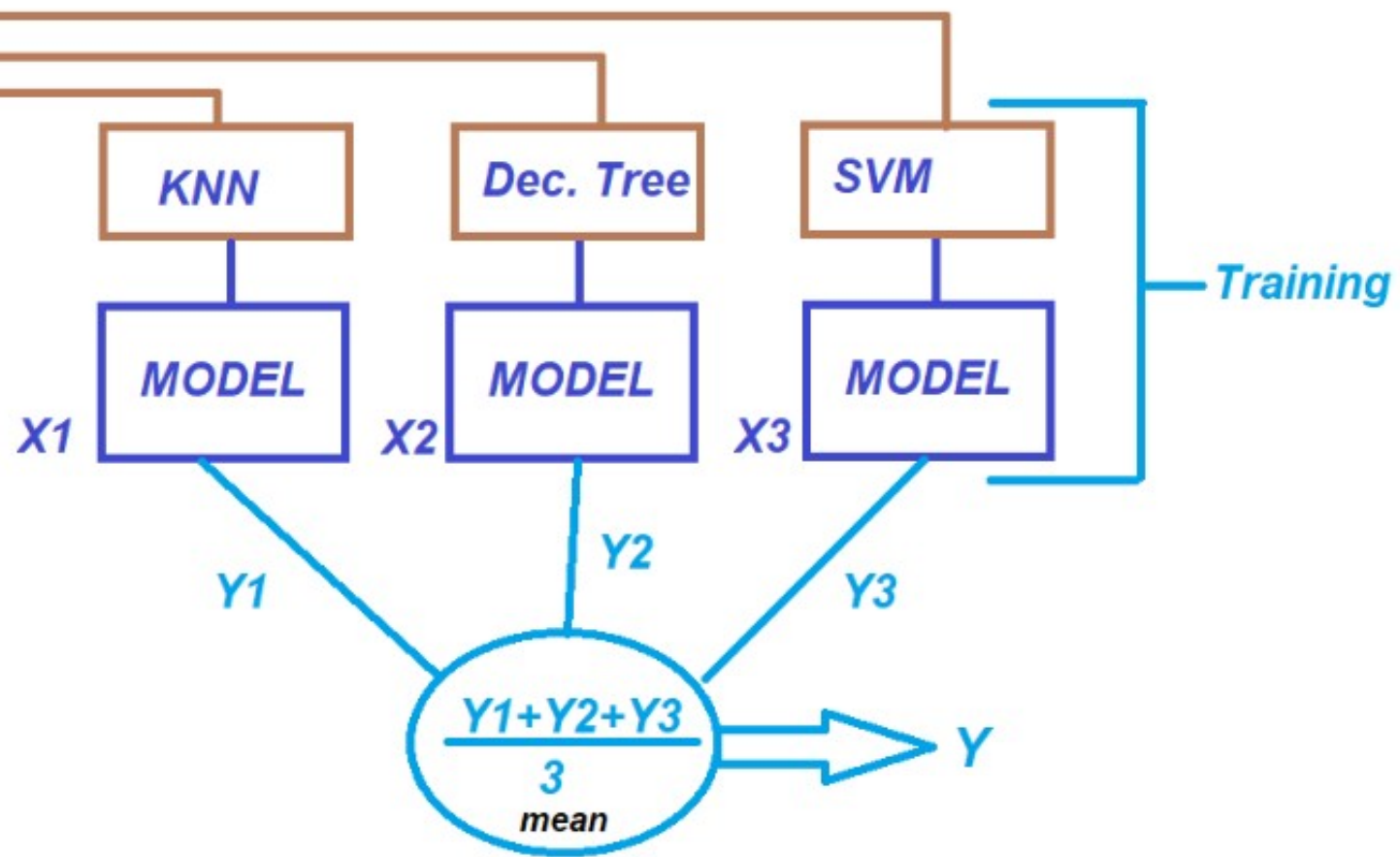
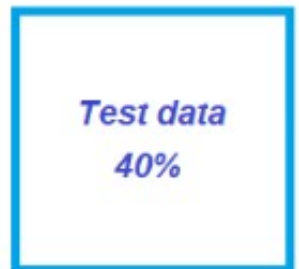
RANDOM FOREST

- Random forest is a ***supervised learning algorithm***.
- It creates decision trees on the given data samples, gets prediction from each tree and selects the best solution by means of voting.
- It is also a pretty good indicator of feature importance.
- ***Random forest algorithm combines multiple decision-trees, resulting in a forest of trees, hence the name Random Forest.***
- ***In the random forest classifier, the higher the number of trees in the forest results in higher accuracy.***

ENSEMBLE TECHNIQUE

- Before we make any big decisions, we ask people's opinions, like our friends, our family members, even our dogs/cats, to prevent us from being biased or irrational.
- Ensemble learning, in general, is a model that makes predictions based on a number of different models.
- By combining individual models, the ensemble model tends to be more flexible and less data-sensitive.

Data



ENSEMBLE TECHNIQUE

- Two most popular ensemble methods are **bagging** and **boosting**.
 - **Bagging:** Training a bunch of individual models in a parallel way. Each model is trained by a random subset of the data
 - **Boosting:** Training a bunch of individual models in a sequential way. Each individual model learns from mistakes made by the previous model

