Machine Learning

Classification and Its Applications

Learning Objectives

By the end of this lesson, you will be able to:

- Comprehend the concept of classification
- List the different types of classification algorithms and their application types
- Choose performance parameters
- Analyze the different types of classifiers

Business Scenario

A bank wants to predict whether a new customer is likely to default on a loan. The bank has a large data set of past customers and their loan repayment history. The bank can use a classification algorithm, such as decision trees or support vector machines (SVM), to classify the new customer as a defaulter or non-defaulter.

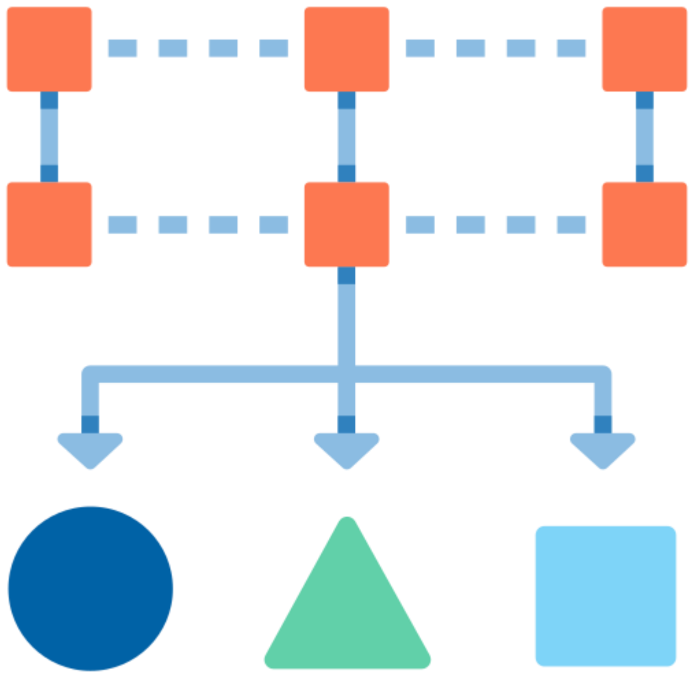
The bank can use performance metrics like accuracy, precision, recall and specificity to evaluate the effectiveness of the classification model. They can also use the gradient descent algorithm to optimize the model and achieve better accuracy.

The K-nearest neighbors (KNN) algorithm can also be used to predict loan defaults based on similar past customers' data. Overall, by utilizing these machine learning algorithms, the bank can make more accurate predictions and reduce the risk of financial losses due to loan defaults.

What Are Classification Algorithms?

Classification

Classification is a supervised learning task in machine learning where the model is trained to predict the class label of a given input data. It looks for the decision boundary, which divides the dataset into different classes.



Classification is an activity that classifies information into subcategories.

This is done by an automated process or a machine.

Classification

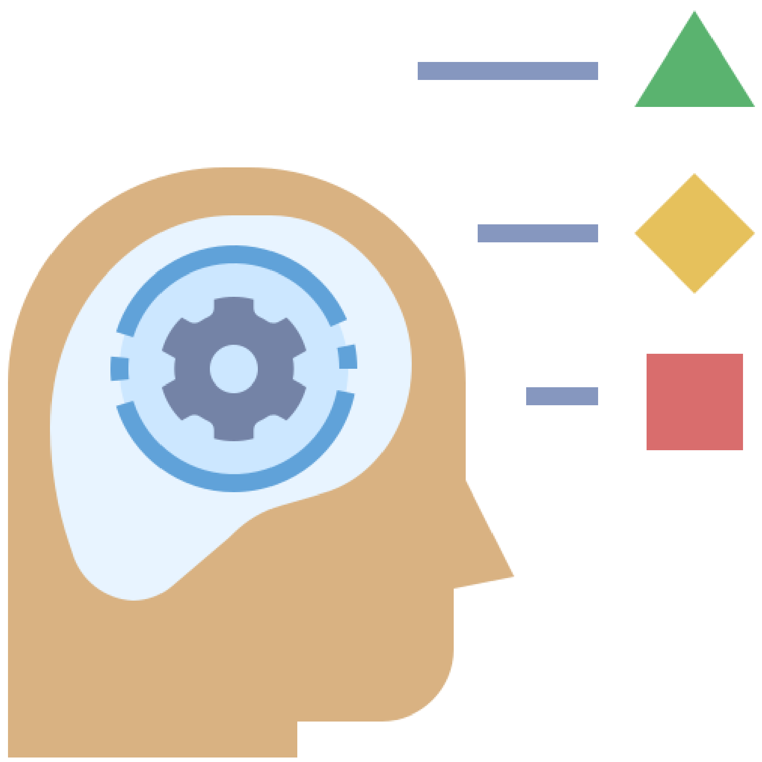
Consider the following examples:

A home security system differentiating between the home resident and an intruder

A machine differentiating between leafy, root and cruciferous vegetables

Classification

Classification involves categorizing training data into different categories (classes) or subpopulations based on their features.



A training set of data contains prior observations whose categories of membership are known.

Classification Algorithms and Classifiers

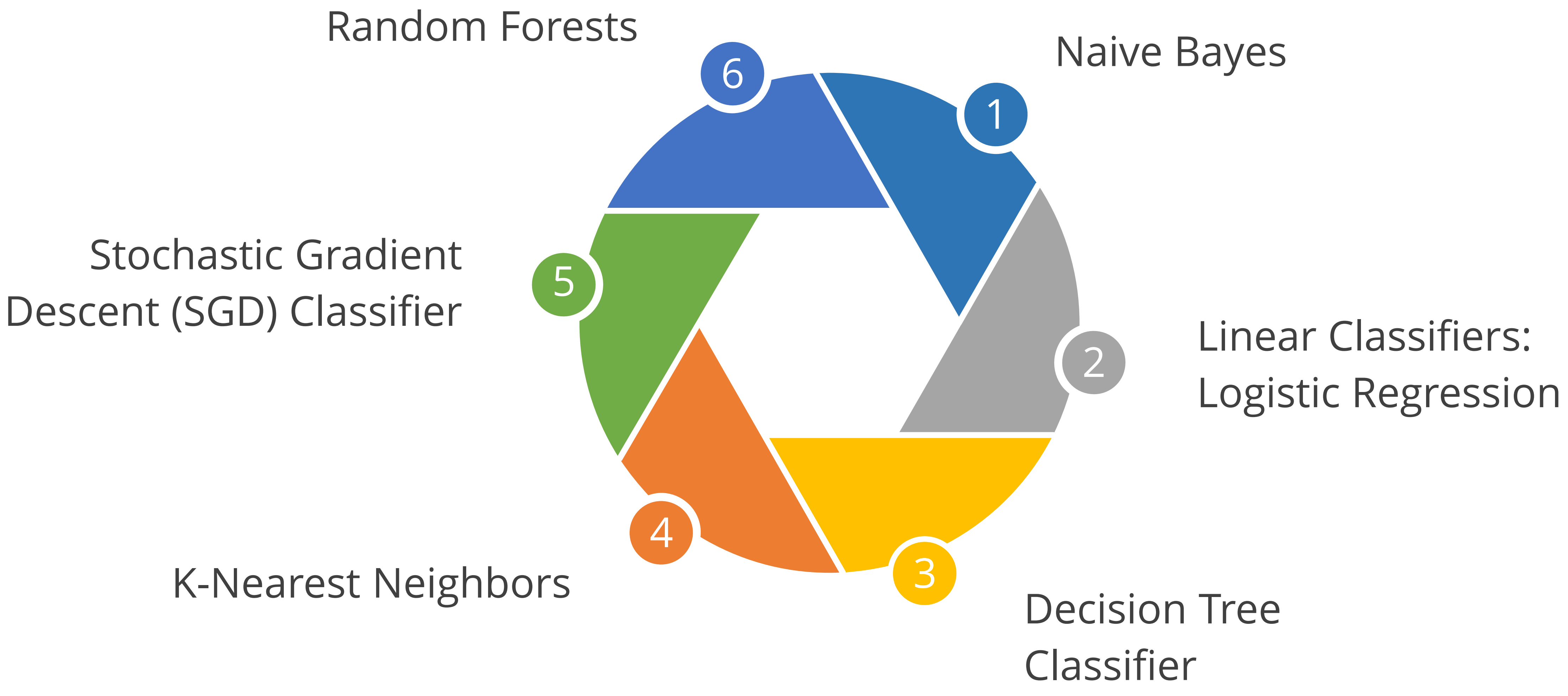
Classification algorithms are used to classify or categorize data into a class or category.

They are also known as classifiers that map input data into a specific category.

Classifiers can operate on both structured and unstructured data.

Classification Algorithms: Types

Some common classification algorithms are:



Key Factors for Selecting a Classification Algorithm

Important factors to consider when choosing and working with classification algorithms are:

1. **Nature of the data:** Number of features, dimensionality, distribution, etc.
2. **Size of the data:** Some algorithms may require a large amount of data to learn and build complex patterns.
3. **Interpretability:** Some algorithms provide interpretable models to help explain the logic behind predictions.

Key Factors for Selecting a Classification Algorithm

Important factors to consider when choosing and working with classification algorithms are:

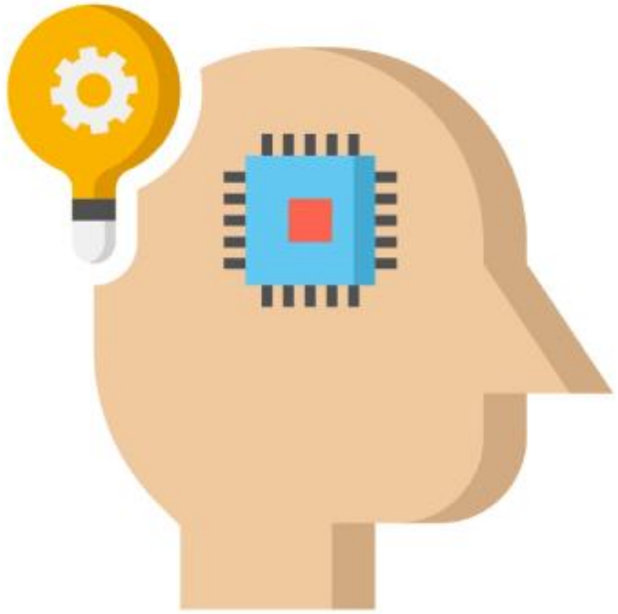
4. **Algorithm assumptions:** It is important to understand specific assumptions about the algorithm.

5. **Computational complexity:** Some algorithms may require heavy computational power. Also, system scalability needs to be considered.

The nature of the input data refers to the attributes of the data, based on which the classification is done.

Classification Algorithms: Applications

The practical applications of machine learning that uses classification algorithms are:



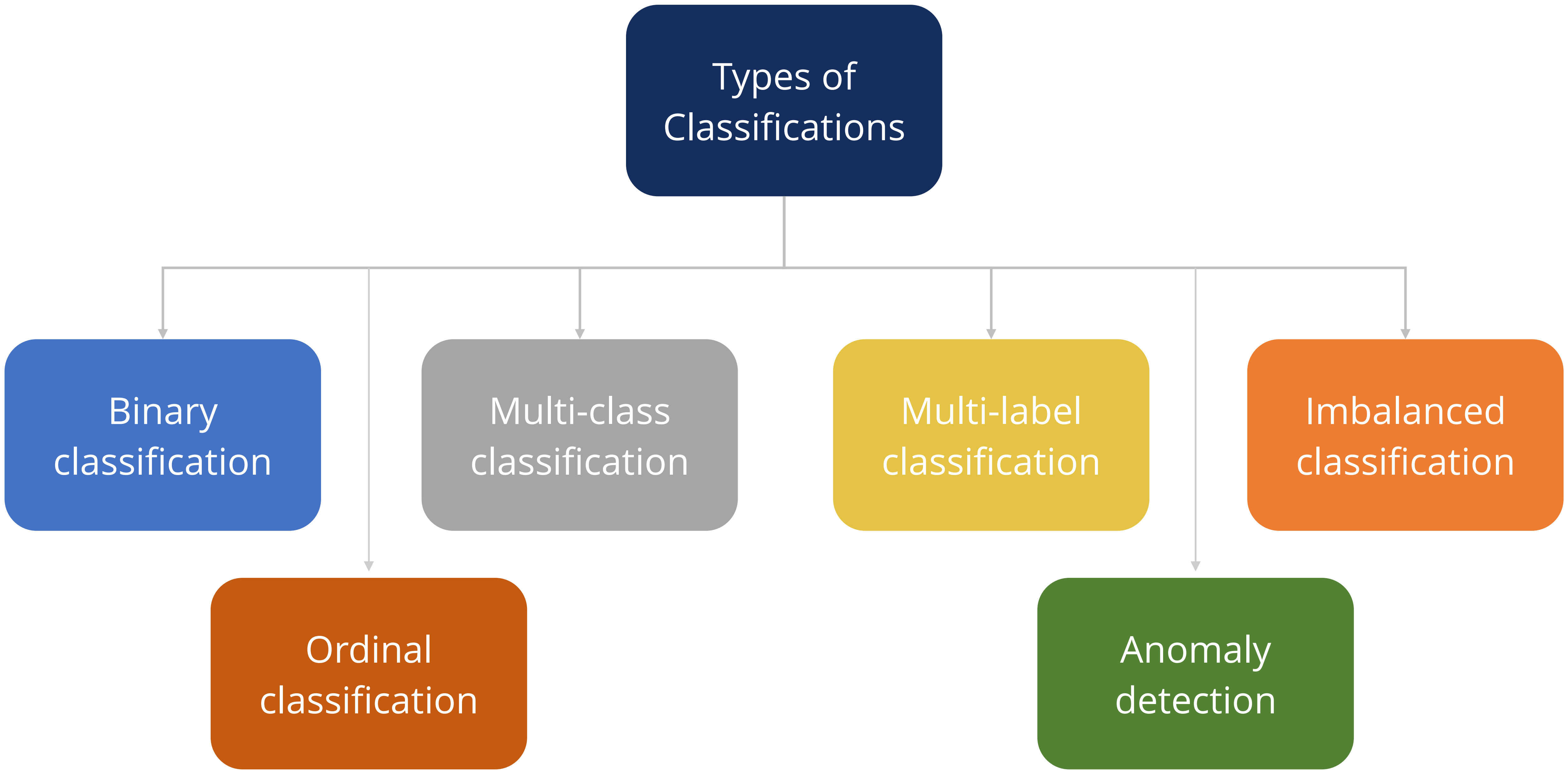
- Detection of health problems
- Classification of the nature of certain diseases like cancer
- Recognition of face and speech patterns
- Navigation of self-driving cars

Businesses can use machine learning algorithms to reduce bias in outcomes and improve quality of products while rapidly scaling the business.

Types of Classification

Types of Classifications

Algorithms can be classified into the following types:



Binary Classification

Binary classification divides the instances into two classes based on the classification rule.

Traditional state (0)	Abnormal state (1)
It is assigned the class label zero.	It is assigned the class label one.

Example

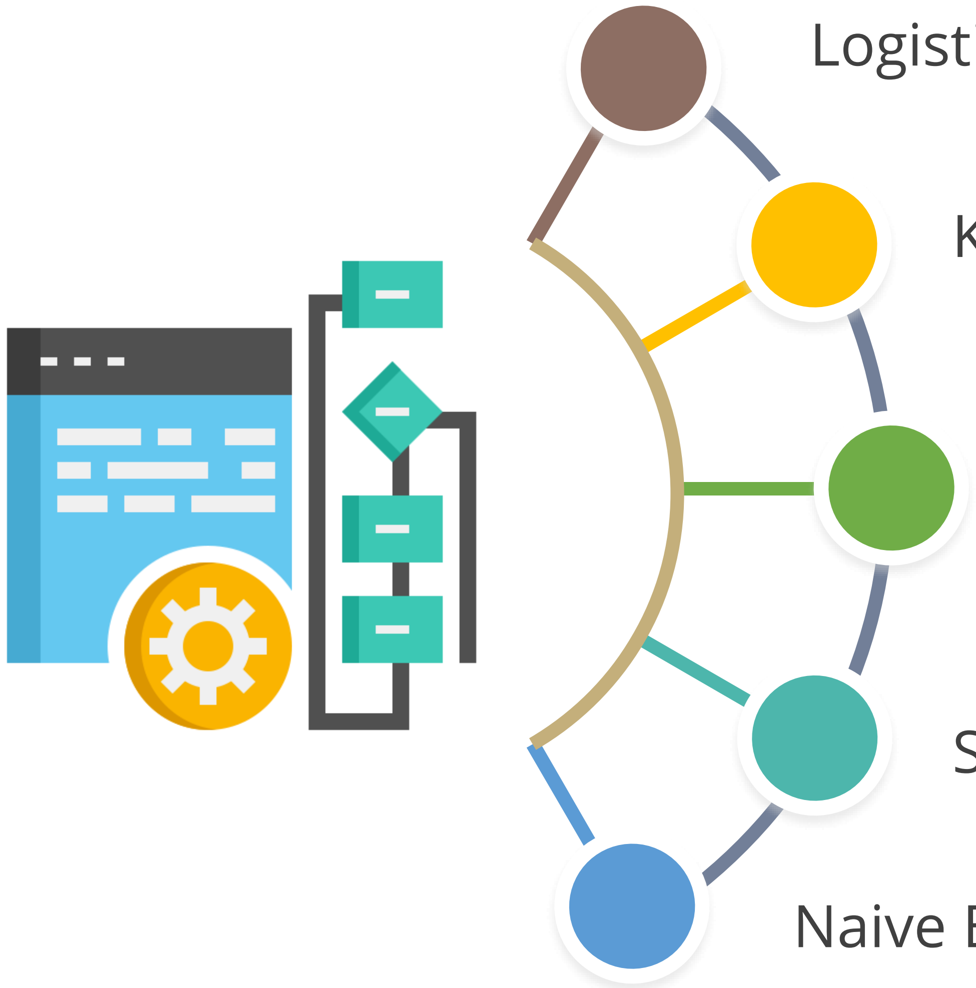
An apple being juicy and fresh

Example

A rotten apple

Binary Classification

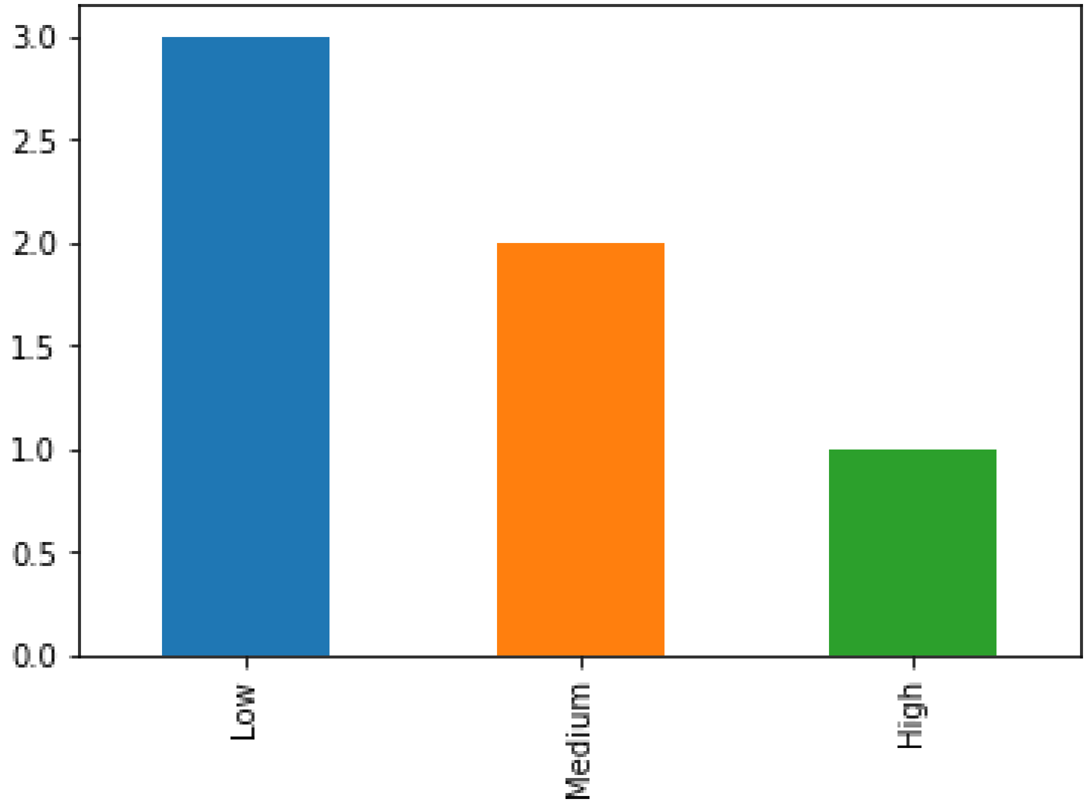
Some of the popular algorithms used for binary classification are:



- Logistic Regression
- K-Nearest Neighbors
- Decision Trees
- Support Vector Machine
- Naive Bayes

Ordinal Classification

Ordinal classification is a type of classification task where the target variable has ordered categories or levels.

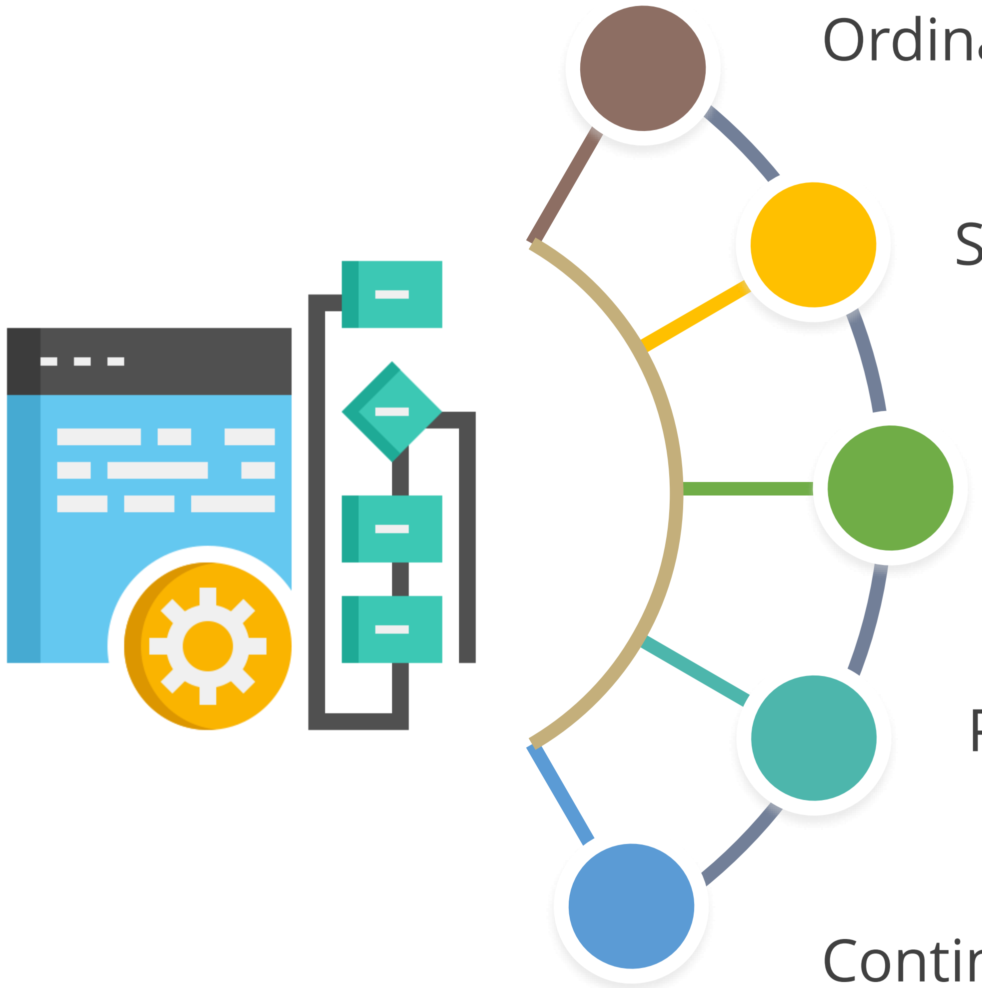


Example

Example: Predicting the student performance levels (low, medium, high) based on study hours, previous test scores, and participation level in a class

Ordinal Classification

Some of the popular algorithms used for ordinal classification are:



- Ordinal Logistic Regression
- Support Vector Ordinal Regression (SVOR)
- Proportional Odds Model
- Partial Proportional Odds Model
- Continuation Ratio Model

Multi-Class Classification

It classifies instances into more than two classes.

The classification tasks can have only one label assigned to them and can be used for cases such as classification of plants into different species and optical character recognition.

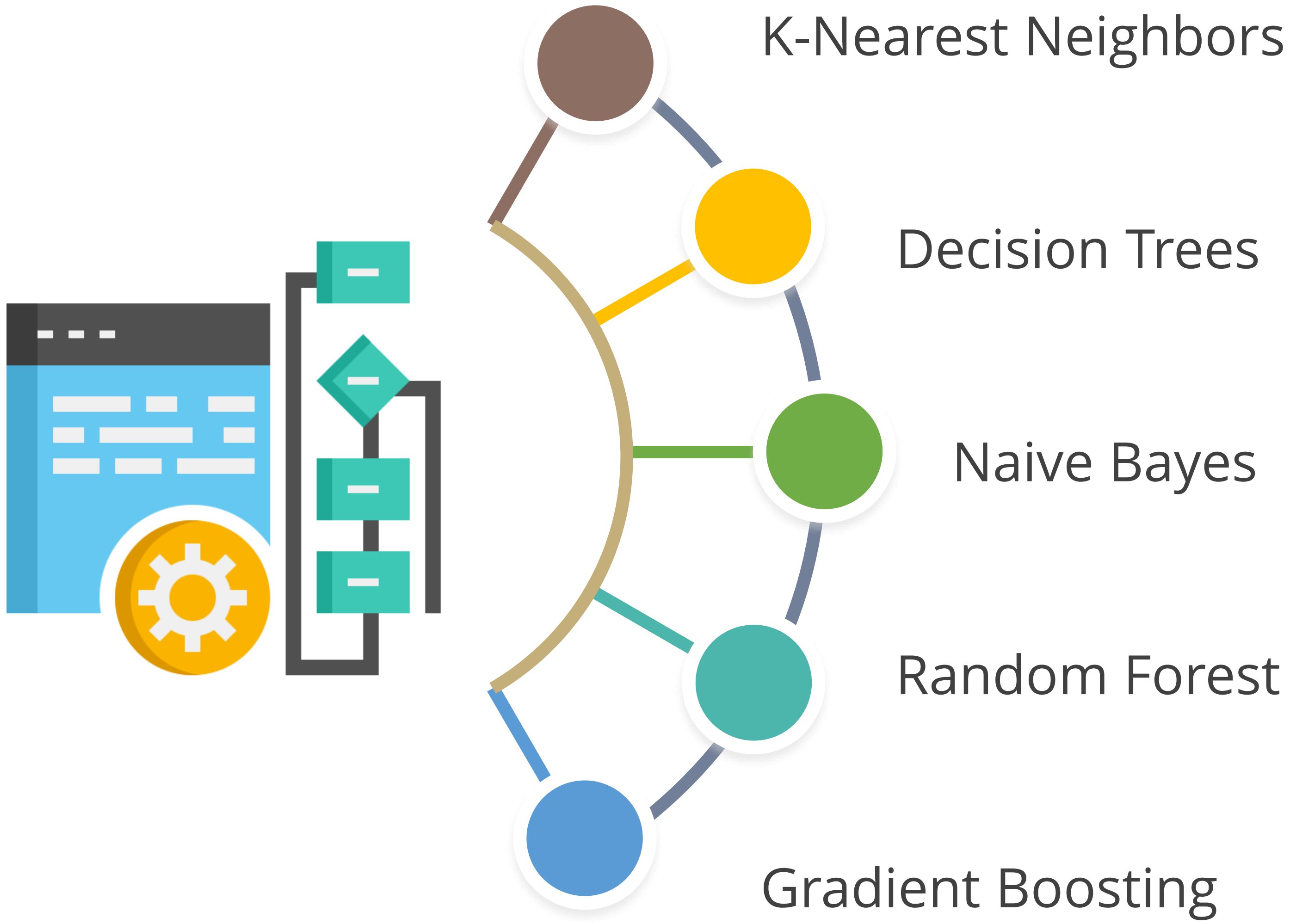
Example

A tree can be a banyan tree or a palm tree but not both.

Multi-class classification classifies the outcome as belonging to one among all the range of known classes.

Multi-Class Classification

Some of the popular algorithms used for multi-class classification are:



Multi-Label Classification

It allows us to classify datasets with more than one target variable.

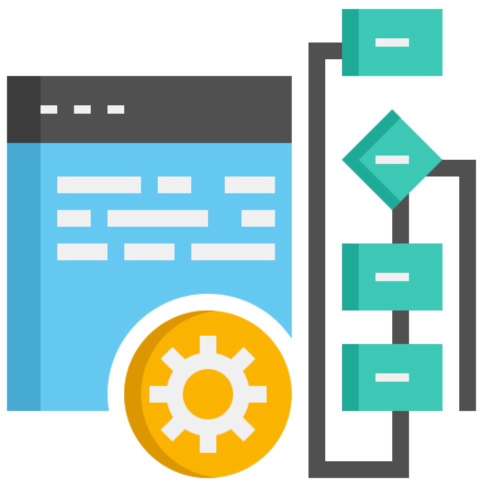
Example

When predicting a given movie's category, it may belong to horror, romance, adventure, action or all simultaneously. Here, there are multi-labels that can be assigned to a given movie.

There are several labels for the outputs for a given prediction, and when making predictions, a given input may belong to more than one label.

Multi-Label Classification

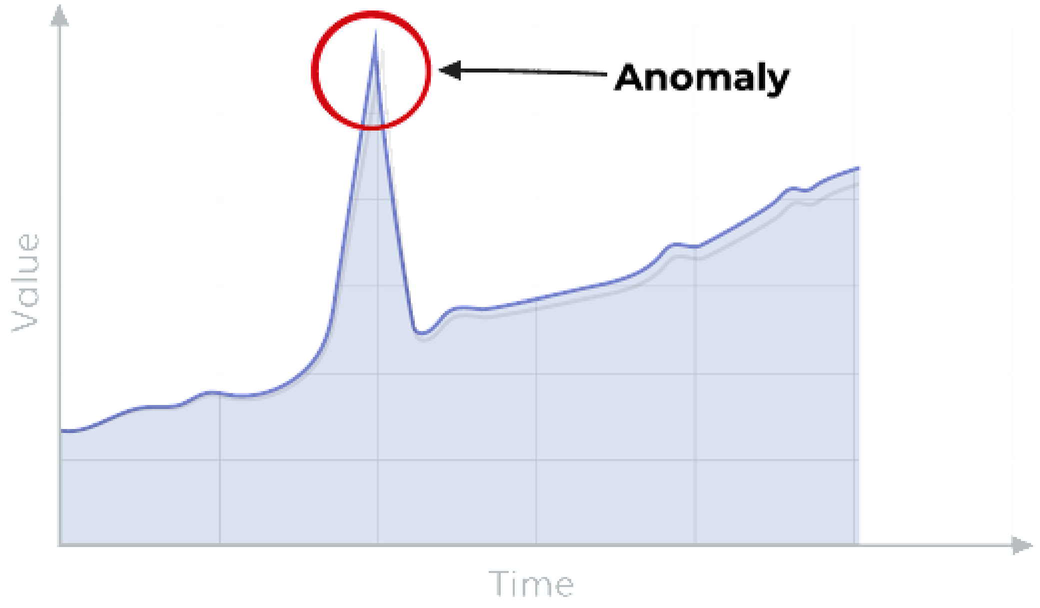
The commonly used algorithms are:



- Multi-Label Decision Trees
- Multi-Label Random Forests
- Multi-Label Gradient Boosting

Anomaly Detection

It is a specialized type of classification algorithm that focuses on identifying rare or unusual instances in data.

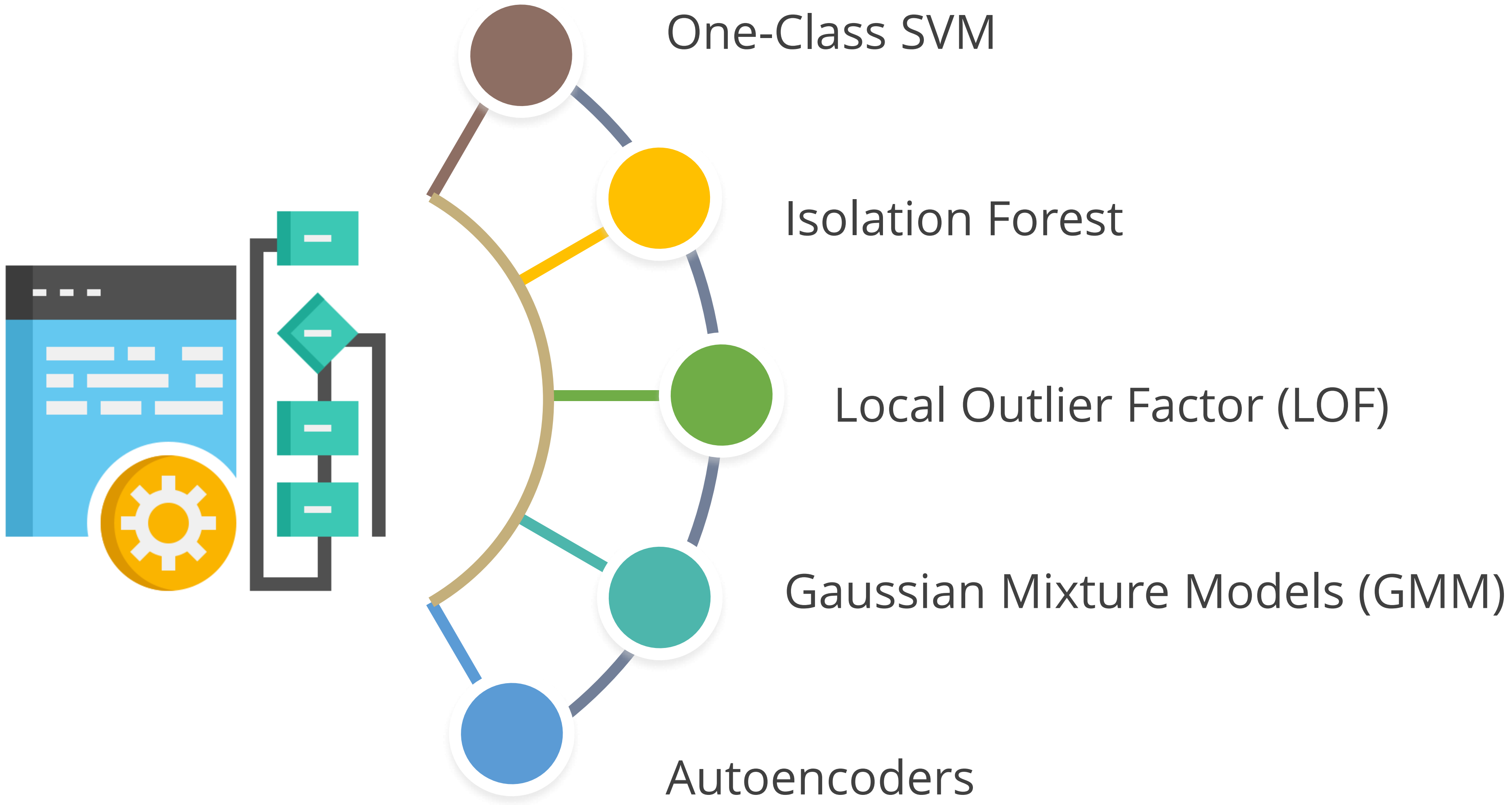


Example

Example: Anomaly detection in credit card fraud detection, where the algorithm classifies transactions as normal or fraudulent based on features like transaction amount, location, and spending patterns, aiming to identify unusual transactions deviating from the expected behavior

Anomaly Detection

Some of the popular algorithms used for anomaly detection are:



Imbalance Classification

Imbalance classification involves unequal distribution of classes.

The imbalance in the class distribution may vary, but a severe imbalance is more challenging to model and may require specialized techniques like:

Random undersampling

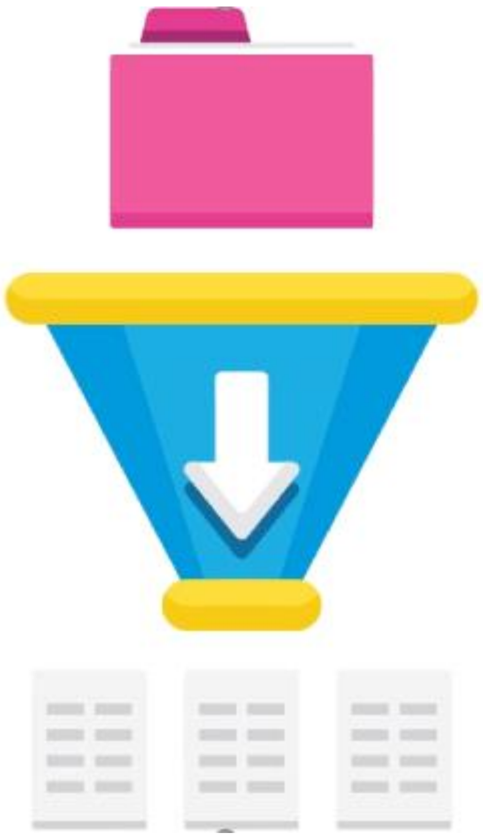
SMOTE oversampling

Many real-world classification problems, such as fraud detection, spam detection and churn prediction, have an imbalanced class distribution.

Application Types and Selection of Performance Parameters

Performance Metrics

Once the type of classification is decided and classifiers are applied on input data, the resulting model output is available with a probability or a category.



Certain metrics are used to determine the effectiveness of the model.

Performance Metrics

Features are chosen while applying classifiers to the data set under consideration.

Metrics are used to monitor and measure model performance on a validation set.

These provide feedback on whether the approach is working or not.

One way to evaluate the performance of a classifier is to look at the confusion matrix.

Confusion Matrix

It is an intuitive and a relatively simple metric that can be effectively used to assess the correctness and accuracy of a model.

Actual

		Positive (1)	Negative (0)
Predicted	Positive (1)	TP	FP
	Negative (0)	FN	TN

It is employed in scenarios where classification tasks involve categorizing data into multiple distinct classes or categories.

Confusion Matrix: Example

Consider the example of a patient being tested for the COVID-19 virus.

		Actual	
		Positive (Covid)	Negative (No Covid-19)
Predicted	Positive (Covid-19)	TP (120)	FP (30)
	Negative (N0 Covid-19)	FN (20)	TN (850)

The provided confusion matrix offers a summary of the effectiveness of COVID-19 tests that identify individuals with the virus and those without it.

Confusion Matrix

Let us compare the patient's test result to the terms used in the confusion matrix.

True positive (TP)	The case detected to be positive is classified correctly as being positive. (The result from the test and the condition of the patient match.)
True negative (TN)	The case detected to be negative is classified correctly as being negative. (The result from the test and the condition of the patient match.)
False positive (FP)	The case is actually negative but was falsely classified as positive. (The result from the test and the condition of the patient do not match.)
False negative (FN)	The case is actually positive but was falsely classified as negative. (The result from the test and the condition of the patient do not match.)

Confusion Matrix

In an ideal scenario, the model classifies correctly and predicts zero false positives and zero false negatives.

		Actual	
		Positive (1)	Negative (0)
Predicted	Positive (1)	TP	FP
	Negative (0)	FN	TN

No model is 100% accurate.

Confusion Matrix

A confusion matrix can be used to evaluate the performance of a model using the following four key metrics:

- Accuracy
- Precision
- Recall or sensitivity
- Specificity

Accuracy

It is the number of correct predictions made by the model divided by the total number of predictions made.

		Actual	
		Positive (1)	Negative (0)
Predicted	Positive (1)	TP	FP
	Negative (0)	FN	TN

Accuracy

$$\frac{TP + TN}{TP + FP + FN + TN}$$

This tells how often the model made correct predictions with respect to the example of the COVID-19 test result.

Precision

Precision is a performance metrics used for pattern recognition and classification in machine learning. It is a good metric to use when the cost of a false positive is high.

		Actual	
		Positive (1)	Negative (0)
Predicted	Positive (1)	TP	FP
	Negative (0)	FN	TN

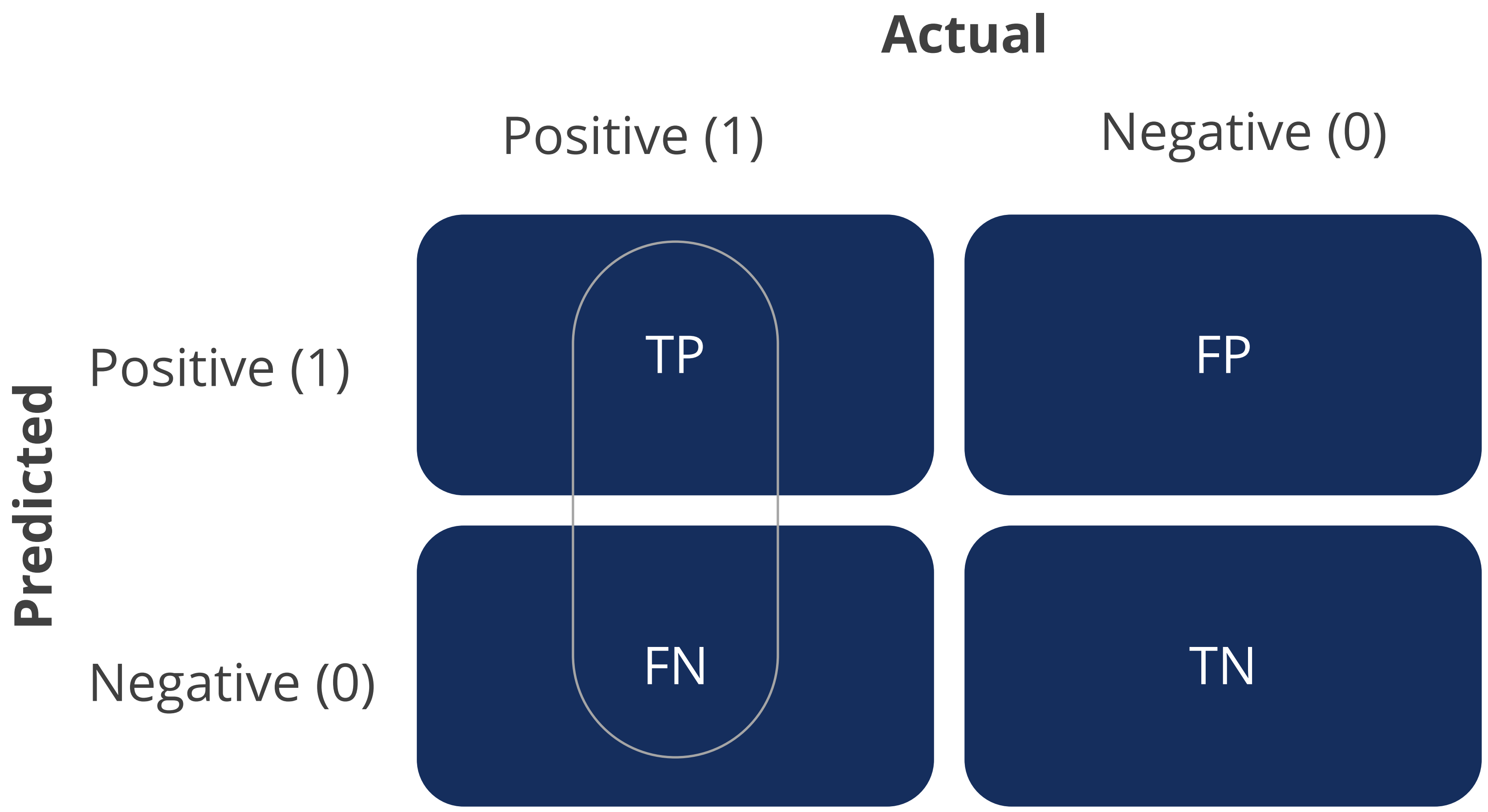
Precision

$$\frac{TP}{TP + FP}$$

In the example, the cost to the patient being falsely detected as being COVID-19 positive is high, resulting in medical costs and possible hospitalization.

Recall or Sensitivity

Recall measures the model's ability to detect positive samples. The higher the recall, the more positive samples detected. It is a very useful metric when the cost of a false negative is high.



Recall

$$\frac{TP}{TP + FN}$$

In the example, if the patient actually has COVID-19 and the model predicts a false outcome, the cost to the patient is high.

Specificity

Specificity is a metric that points to how many cases that were actually COVID-19 negative were correctly reported as negative.

Actual

Predicted	Positive (1)	Negative (0)
Positive (1)	TP	FP
Negative (0)	FN	TN

Specificity

$$\frac{TN}{TN + FP}$$

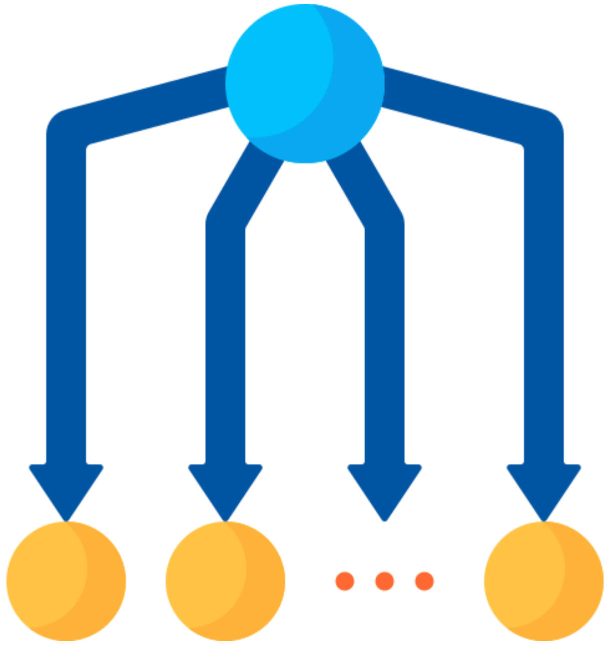
Naive Bayes

Discussion: Naive Bayes Classifier

- Can missing values impact the Naive Bayes algorithm?
- Discuss some use cases where a Naive Bayes classifier can be used.

Naive Bayes

The Naive Bayes classifier is a machine learning model that segregates different objects on the basis of certain features of variables.



Example

The probability of the price of a house being high can be assessed with knowledge about the neighborhood.

The Naive Bayes classifiers are based on the Bayes theorem, and it predicts the probable occurrence of an event based on the prior knowledge of other associated events.

Principle of Contingent Probability

Bayes theorem works on the principle of contingent probability.

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

P(A) is the probability of the hypothesis H being true

P(B) is the probability of evidence being true

P(B|A) is the probability of the evidence given that the hypothesis is true

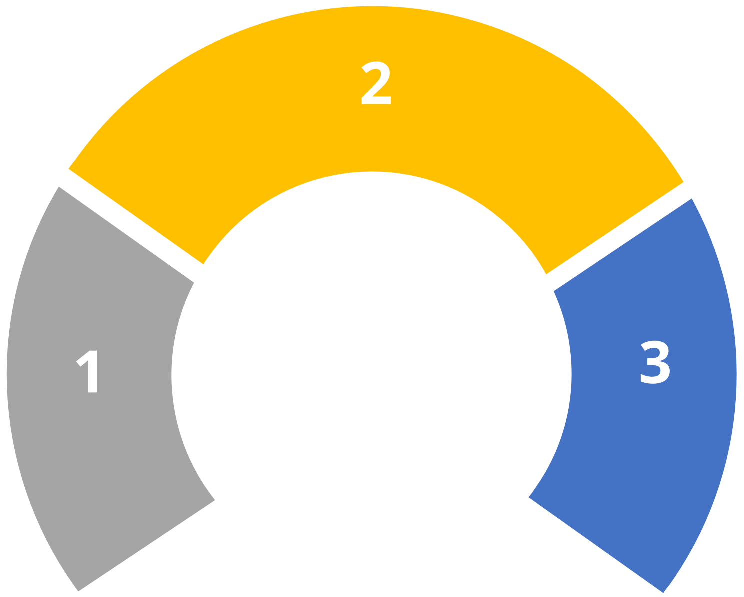
P(A|B) is the probability of the hypothesis given that the evidence is true

It is a measure of the probability that something will happen given that something else has already occurred.

Types of Naive Bayes Algorithms

The different types of Naive Bayes algorithms are:

Multinomial Naive Bayes



Gaussian Naive Bayes

Bernoulli Naive Bayes

Applications of Naive Bayes

It identifies the right fit for a job by classifying the skills mentioned in the resumes of candidates.

It classifies content available in the publishing field based on the availability of the information in different sources, like the press and social media.

It classifies the relevant tags and research papers and helps researchers gain access to the right content.

It classifies legal papers and identifies illegal documents in any field.

Applications of Naive Bayes

It classifies speeches and choice of words of political candidates and determines the candidates' mentality. It can also check for words or sentences related to racial abuse.

It checks if there are any degrading sentences or words related to colorism in ad dialogues.

It checks for gender discrimination in hiring practices.

Discussion: Naive Bayes Classifier

- Can missing values impact the Naive Bayes algorithm?

Answer: Yes, missing values can impact the Naive Bayes algorithm by violating its assumption of feature independence and potentially leading to biased probability estimates.

- Discuss some use cases where a Naive Bayes classifier can be used.

Answer: Naive Bayes classifier can be used to:

- classify legal papers and identify illegal documents
- classify content available in the publishing field
- classify the relevant tags and research papers

Assisted Practices

Let’s understand the topic below using Jupyter Notebook.

- 5.06_Applying Naive Bayes Classifier

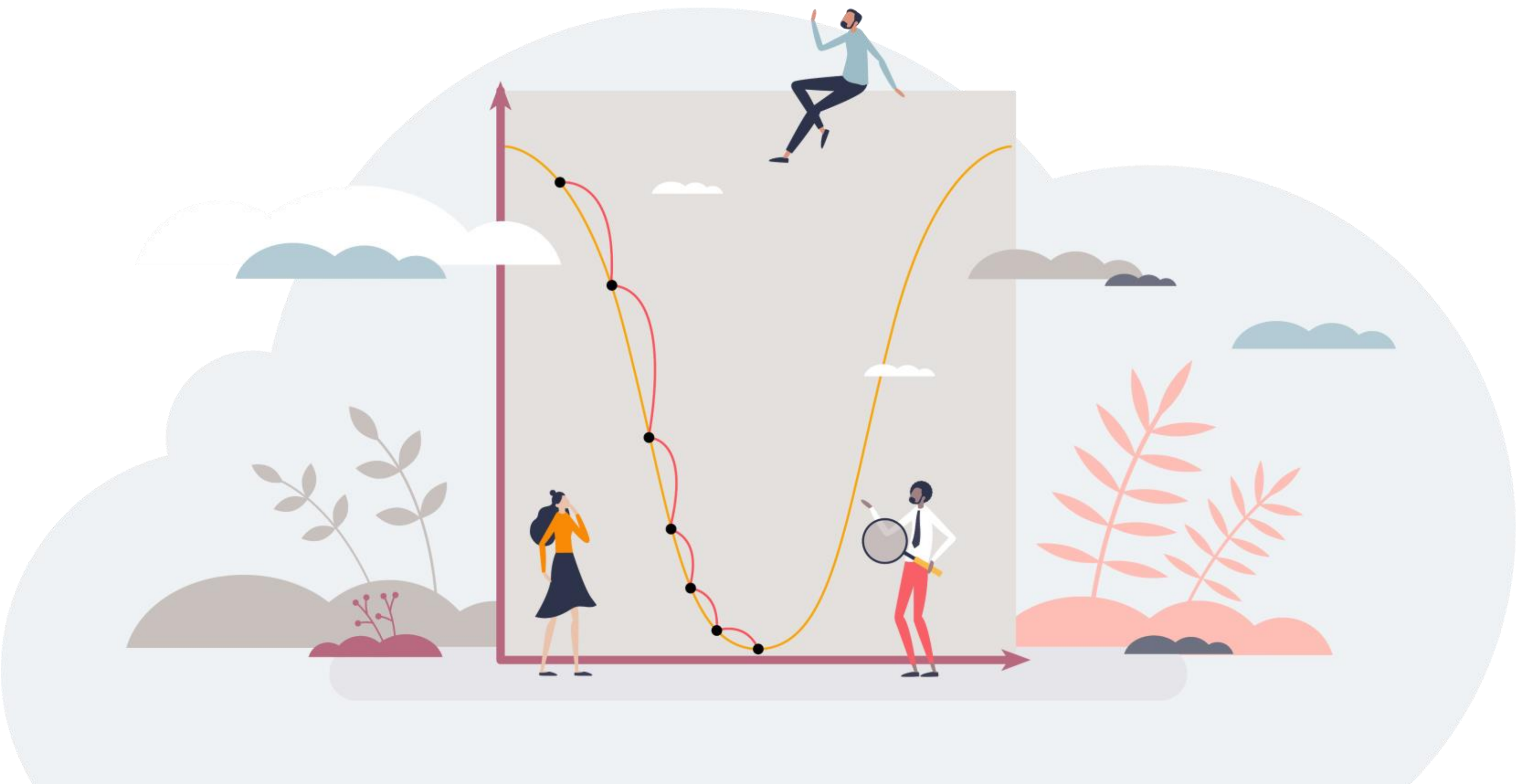
Note: Please download the pdf files for each topic mentioned above from the Reference Material section.

ASSISTED PRACTICE

Stochastic Gradient Descent

Stochastic Gradient Descent (SGD)

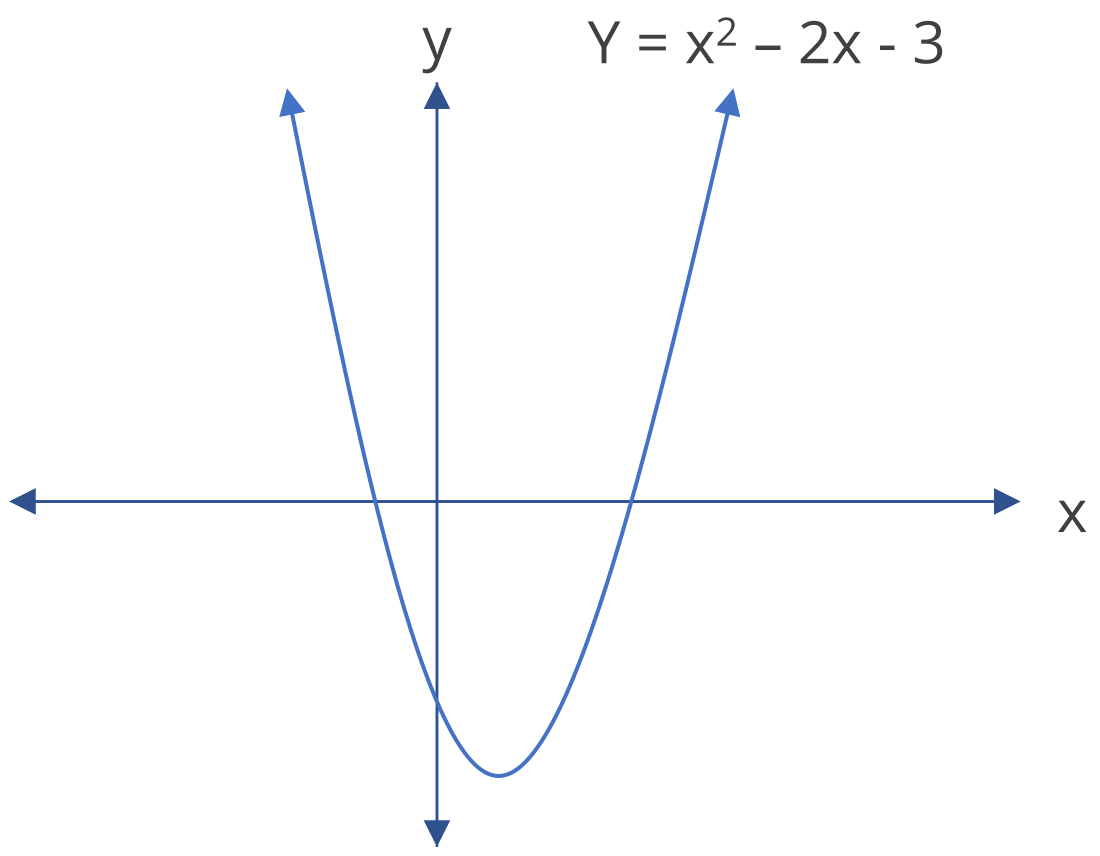
It is a common machine learning algorithm and forms the basis of neural networks.



The word gradient means a slope or slant.

Stochastic Gradient Descent (SGD)

Gradient descent is a descending slope that helps reach the lowest point on a surface.



The lowest point of the parabola occurs when x=1.

The gradient descent algorithm aims to find the value of **x** such that the value of **y** is minimum.

y is the objective function to descend to the lowest point.

It is an iterative algorithm that starts from a random point on a function and then travels down its slope in steps until it reaches the lowest point of that function.

Gradient Descent Algorithm: Steps

The steps involved in a gradient descent algorithm are:

1. Compute the gradient of the function that helps find the slope of the objective function
2. Pick a random initial value for the parameters (in this case, differentiate y with respect to x)
3. Plug in the parameter values and update the function
4. Calculate the step size for each feature: Step size = gradient * learning rate
5. Calculate the new parameters: New params = old params - step size
6. Repeat steps 3 to 5 till the gradient is almost zero

Learning Rate

The learning rate is a flexible parameter that influences the algorithm convergence towards the minimum.

The larger the learning rates, the larger the steps down the slope.

This may lead to the minimum getting missed.

It is thus advisable to use small learning rates.

Stochastic Gradient Descent (SGD)

Researchers have found that the gradient descent algorithm becomes very slow on large data sets.

It randomly picks one data point in the whole data set at each iteration, reducing the number of computations.

Example

A set of 10,000 data points and 10 features may take as much as 10 million computations to complete the algorithm.

Advantages and Disadvantages of SGD

Advantages

It offers speed, efficiency and ease of implementation.

Disadvantages

Due to frequent updates, the steps taken toward the minimum may be very noisy, which may lengthen the time taken to reach the minima.

Applying frequent updates on a large dataset is computationally expensive.

Applications of SGD

When companies find it difficult to keep a check on quality, machine learning helps to automate the process.

It helps tackle low performance issues in the model created.

It helps speed up convergence in huge data sets.

It is used in financial modeling to reduce inaccuracies in predictions.

Assisted Practices

Let’s understand the topic below using Jupyter Notebook.

- 5.08_Applying Stochastic Gradient Descent

Note: Please download the pdf files for each topic mentioned above from the Reference Material section.

ASSISTED PRACTICE

K-Nearest Neighbors

K-Nearest Neighbor

K-nearest neighbor (KNN) is a machine learning algorithm that supports the supervised learning technique.

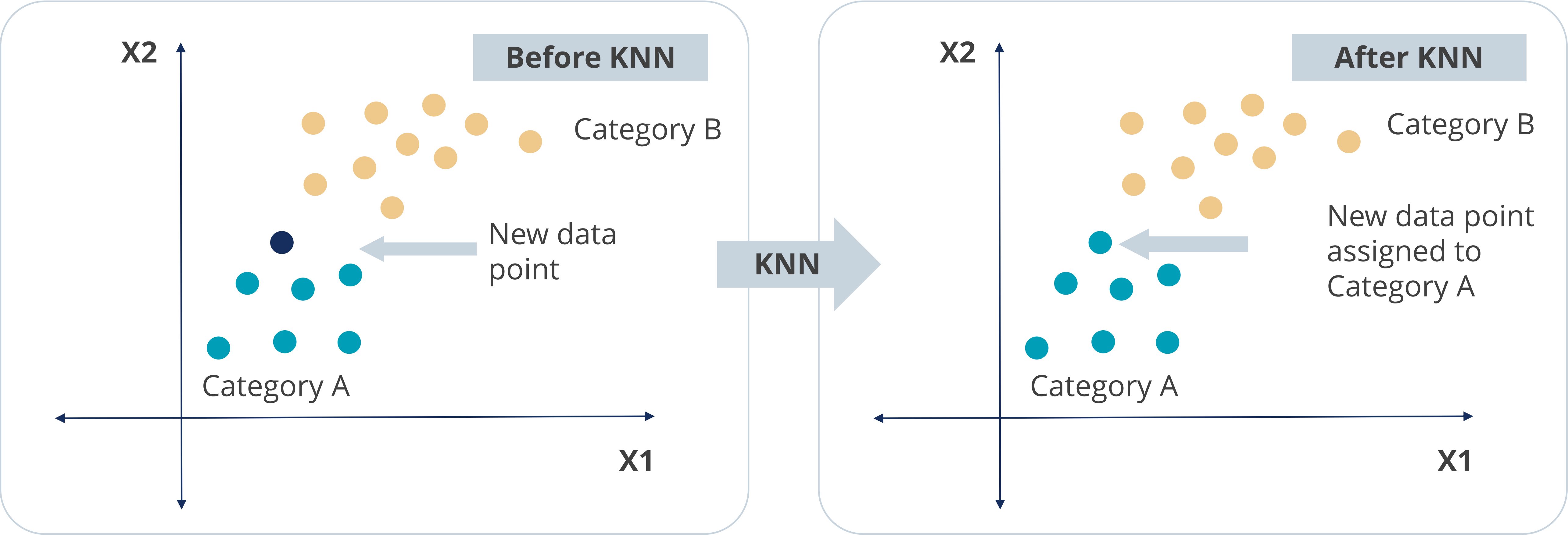
It stores all the available data and classifies data points based on the similarity measures.

It is used for regression and classification problems.

The algorithm assumes that the new and existing instances are similar and places the new instance in the most similar category to the existing ones.

K-Nearest Neighbor

The following diagram represents a before and after KNN scenario.



K-Nearest Neighbor

KNN is a non-parametric algorithm and doesn't make assumptions about the data.

It stores the data set at its training phase.

It assigns new data points to their nearest category that has similar characteristics.

How Does KNN Work?

The following is a step-by-step KNN process:

1. Load the training with test data

2. Choose the value of K, that is the nearest data point (K can be any integer)
Note: The best value for K will depend on the specific dataset and the problem to be solved.

How Does KNN Work?

3 For every point within the test data, do the following step:

3.1 Calculate the space between test data and coaching data with tactics such as Euclidean, Manhattan, or Hamming distance

3.2 Support the space value and sort it in an ascending order

3.3 Choose the highest k rows from the sorted array

3.4 Assign a category to the test point supporting the most frequent class

Advantages of KNN Algorithm

The KNN algorithm has the following advantages:

- Robust even with noisy training data
- Simple to implement
- Simple and effective if the training data is large

Disadvantages of KNN Algorithm

The KNN algorithm has the following disadvantages:

- The worth of K must always be determined, which is quite complex.
- It can be computationally expensive to calculate the space between the info points for all training samples.

Applications of KNN Algorithm

KNN is applied in the manufacturing industry to classify manufactured goods based on:

Size

Grading

Priority

Location

Assisted Practices

Let's understand the topic below using Jupyter Notebook.

- 5.10_Applying K Nearest Neighbors

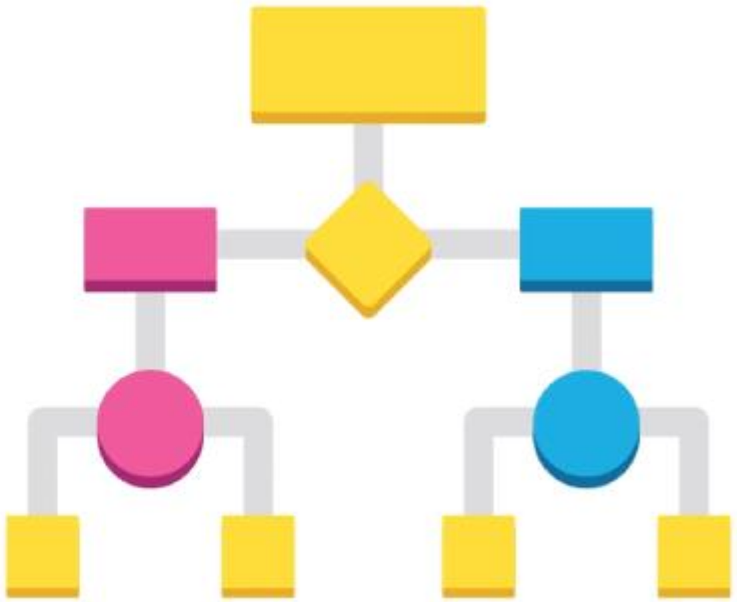
Note: Please download the pdf files for each topic mentioned above from the Reference Material section.

ASSISTED PRACTICE

Decision Tree

Decision Tree (DT)

It is a supervised learning nonparametric method used for classification and regression.



It uses a tree-like structure to explain a business problem.

Decision Tree (DT)

The tree-like structure consists of nodes, branches and leaves.

Nodes

Represent a place where a question is asked, or an attribute is selected

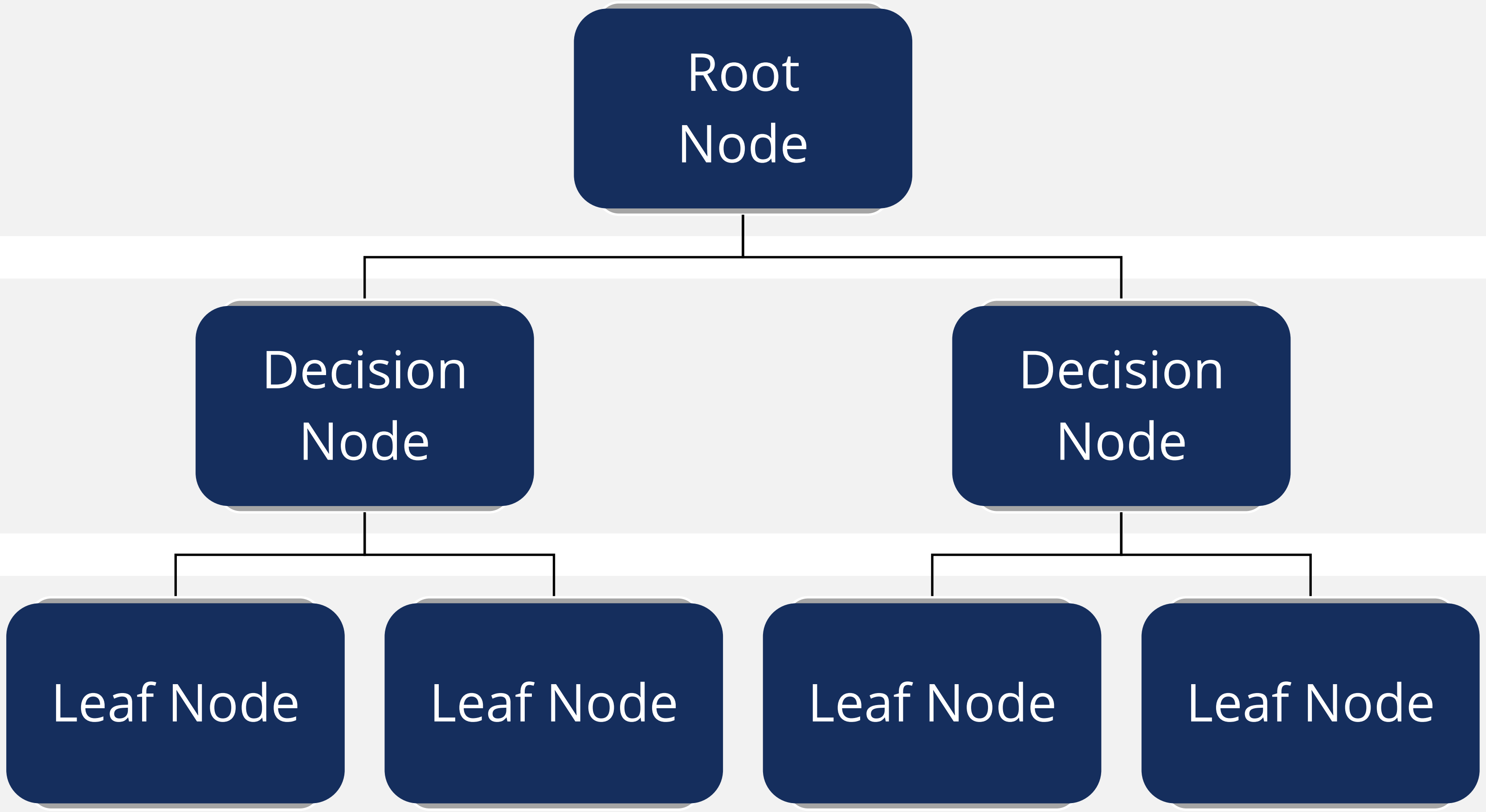
Leaves

Represent the output or class label

Using such a structure and the associated decision rules, the model predicts the value of a target.

Decision Tree (DT)

The following diagram demonstrates the different parts of a decision tree:



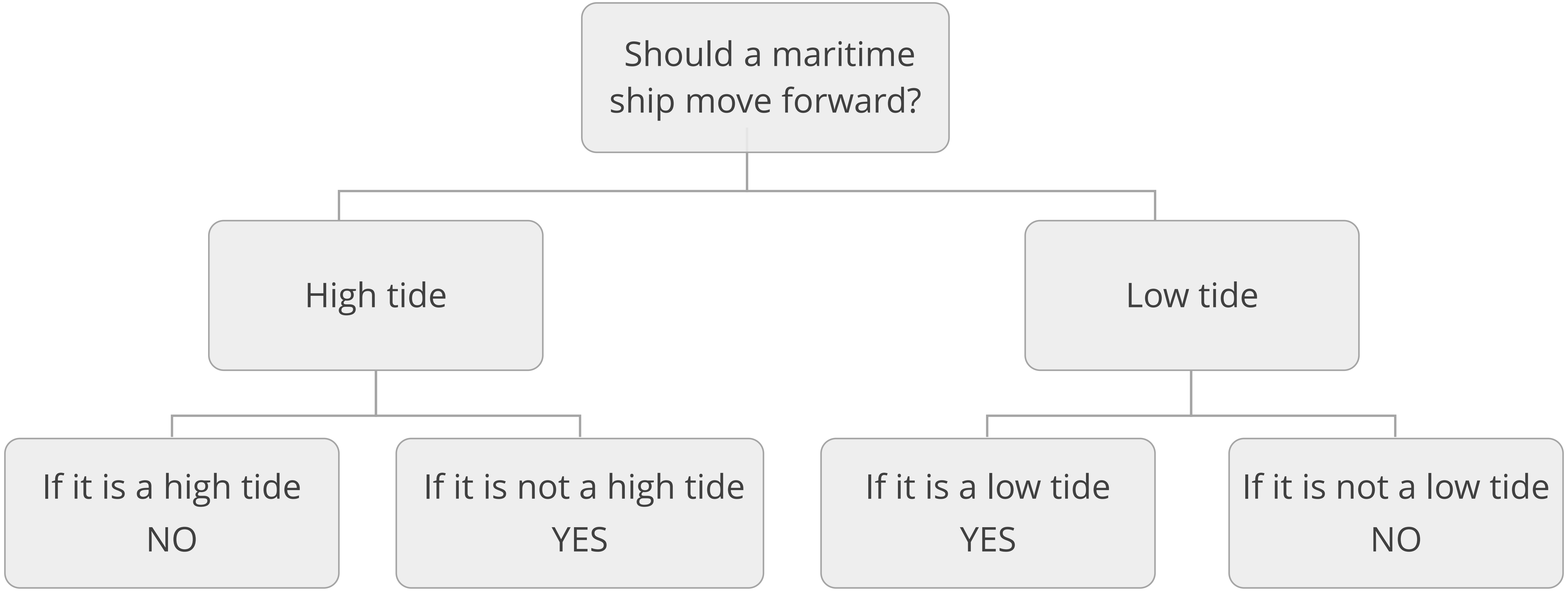
The root node represents a condition that is split into two results.

The decision node continues until the tree has no condition left.

When no further branches arise, it is called a leaf node.

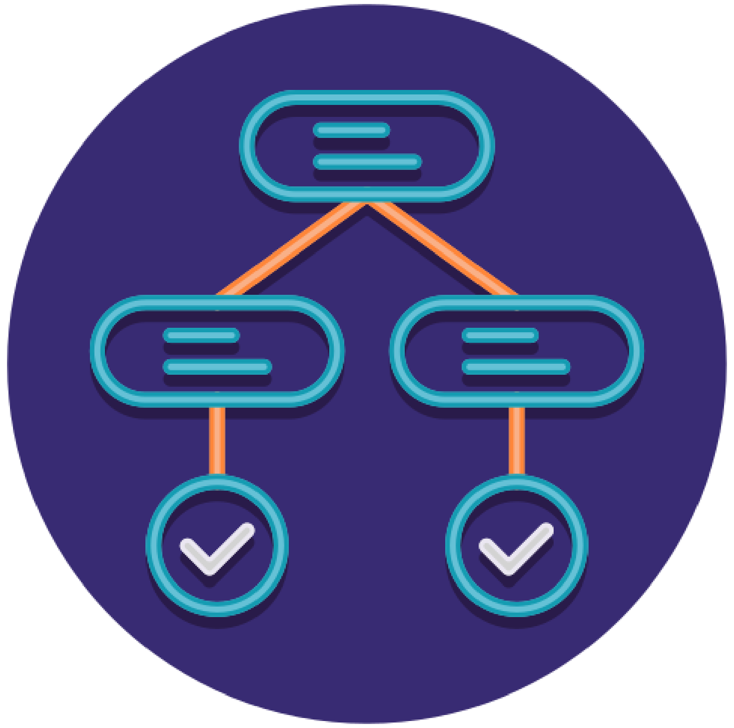
Decision Tree: Example

The following decision tree diagram represents a situation where a maritime ship only moves forward when there is a low tide to save fuel.



Decision Tree

The decision to split into nodes and leaves influences the model's accuracy.



It uses multiple algorithms to separate a node into subnodes.

Division of a parent node increases the homogeneity in subnodes.

Purity of the node is connected to the target variable.

Using such a structure and the associated decision rules, the model predicts the value of a target.

Algorithms Used by Decision Tree

A decision tree uses the following algorithms:

- ID3 (Extension of D3)
- C4.5 (Successor of ID3)
- CART (Classification and regression tree)
- CHAID (Chi-square automatic interaction detection)
- MARS (Multivariate adaptive regression splines)

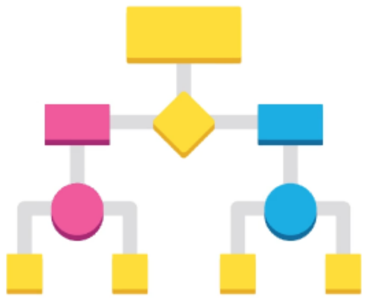
The type of target variable that is selected is an additional input for algorithm selection.

Decision Tree

A decision tree has the following benefits and drawbacks:

Benefits:

- Captures the interactions between features within data
- Easy to interpret
- Does not require normalization of data



Drawbacks:

- Does not consider linear relationships
- Needs more time to train the mode
- Is expensive as the complexity and time taken are greater

Applications of Decision Tree

Decision trees are effective in applications with nonlinear data sets.

Decision trees are used in:

City planning

Law

Engineering

Business applications

Assisted Practices

Let’s understand the topic below using Jupyter Notebook.

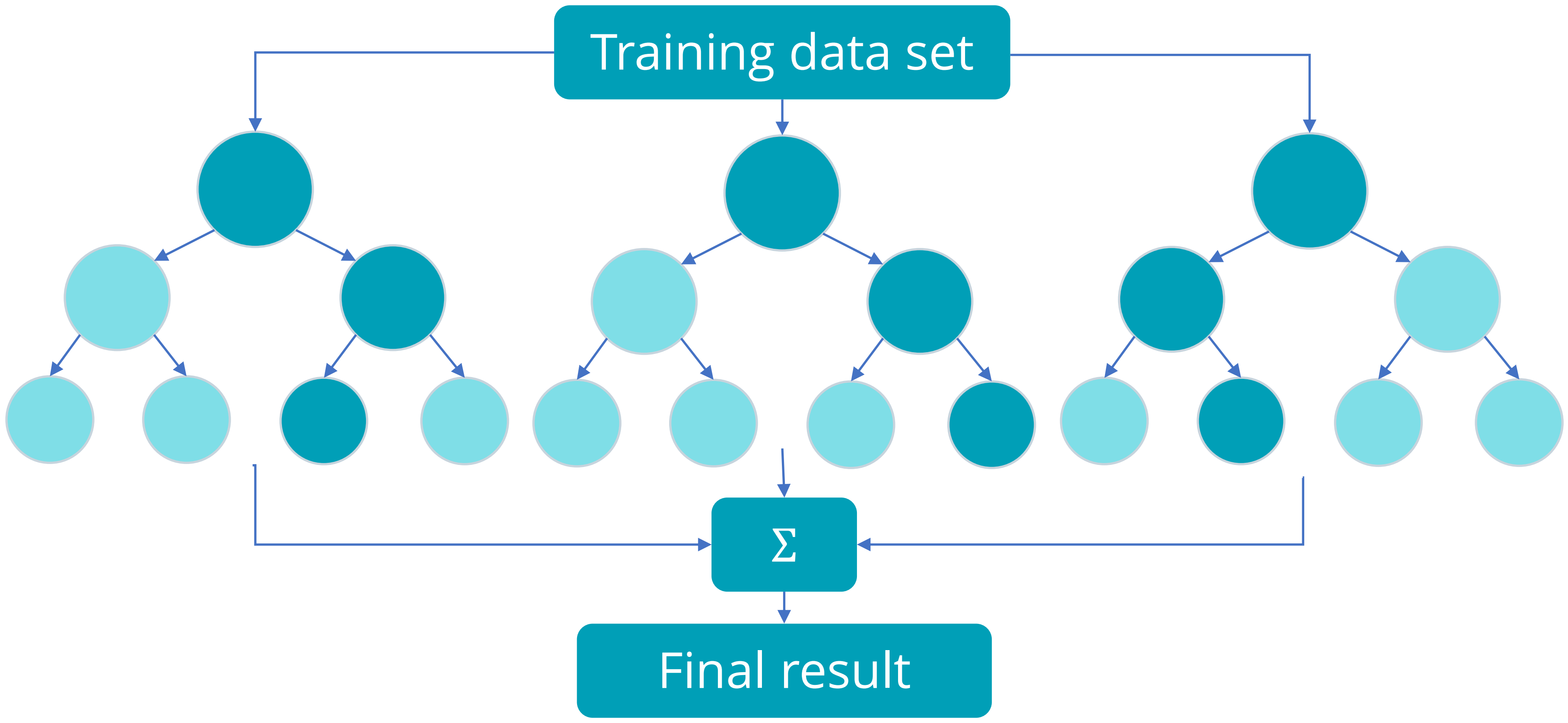
- 5.12_Applying Decision Tree

Note: Please download the pdf files for each topic mentioned above from the Reference Material section.

Random Forest

Random Forest

Random forest is a machine learning technique used to solve regression and classification problems.

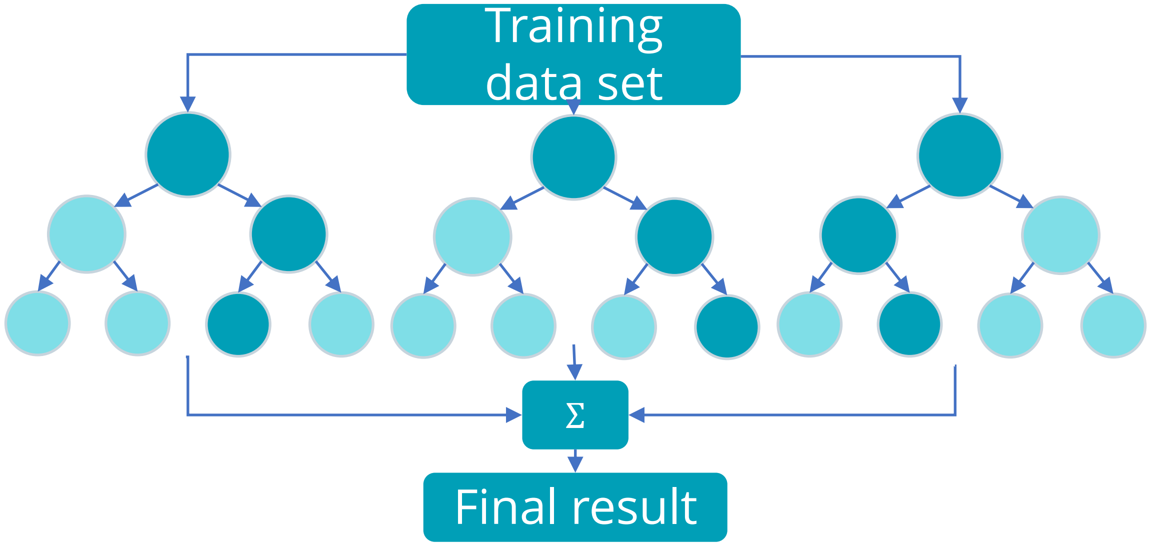


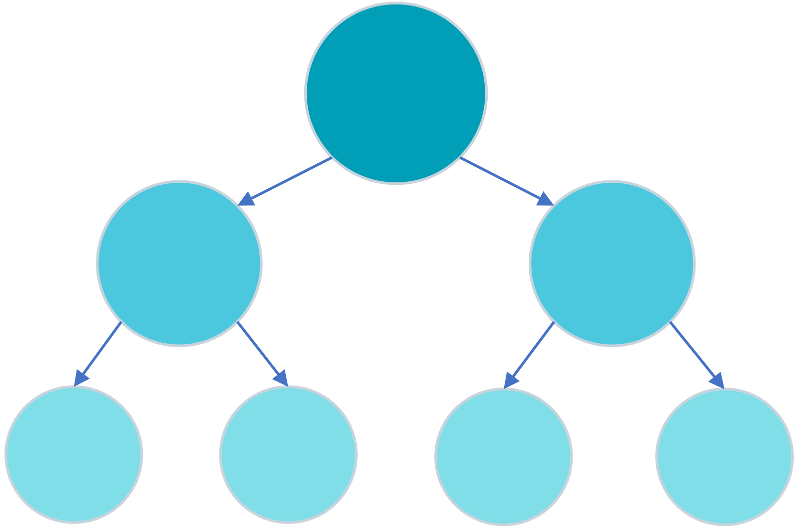
It is an ensemble learning technique which combines multiple classifiers to solve a complex problem.

Random Forest

The random forest algorithm consists of many decision trees.

The key difference between a typical decision tree and a random forest algorithm is that:





Random forests select a subset of the features.

Decision trees consider all the feature splits that are possible.

Advantages of Random Forest

Reduced risk of overfitting: In decision trees, data is fixed in nodes and leaves, so making predictions is easy, but in a random forest it reduces the risk of overfitting.

Provides flexibility: It can handle the data set for classification as well as regression.

Easy to determine: It makes it easy to determine which variable or feature is more important and has the highest contribution towards the model.

Disadvantages of Random Forest

Time-consuming process: The algorithm works best with a large amount of data, but it also slows down the processing.

Requires more resources: It doesn't work with a small data set.

More complex: Interpreting the prediction of a forest can get complex.

Applications of Random Forest

The algorithm is used across multiple industries.

In finance, it reduces the time required to manage a huge data set.

It evaluates prices in stock markets or in banks to process a loan.

It identifies the patterns of employee attrition.

Assisted Practices

Let’s understand the topic below using Jupyter Notebook.

- 5.14_Applying Random Forest

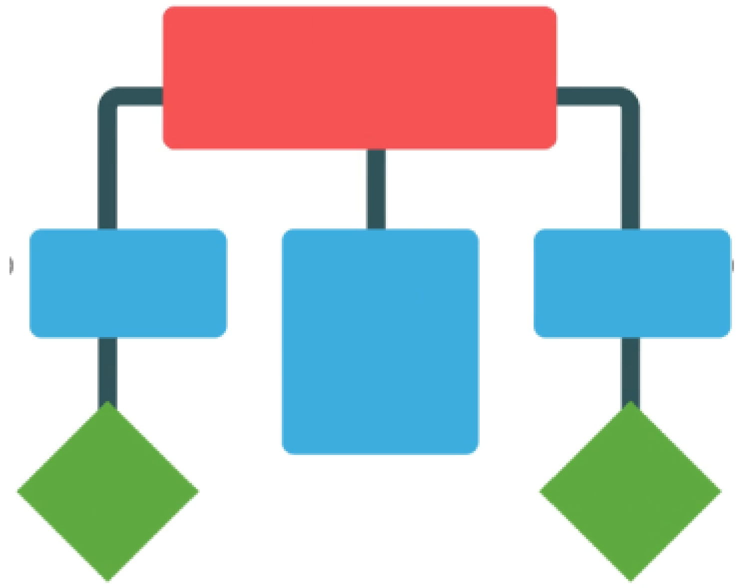
Note: Please download the pdf files for each topic mentioned above from the Reference Material section.

ASSISTED PRACTICE

Boruta

Boruta

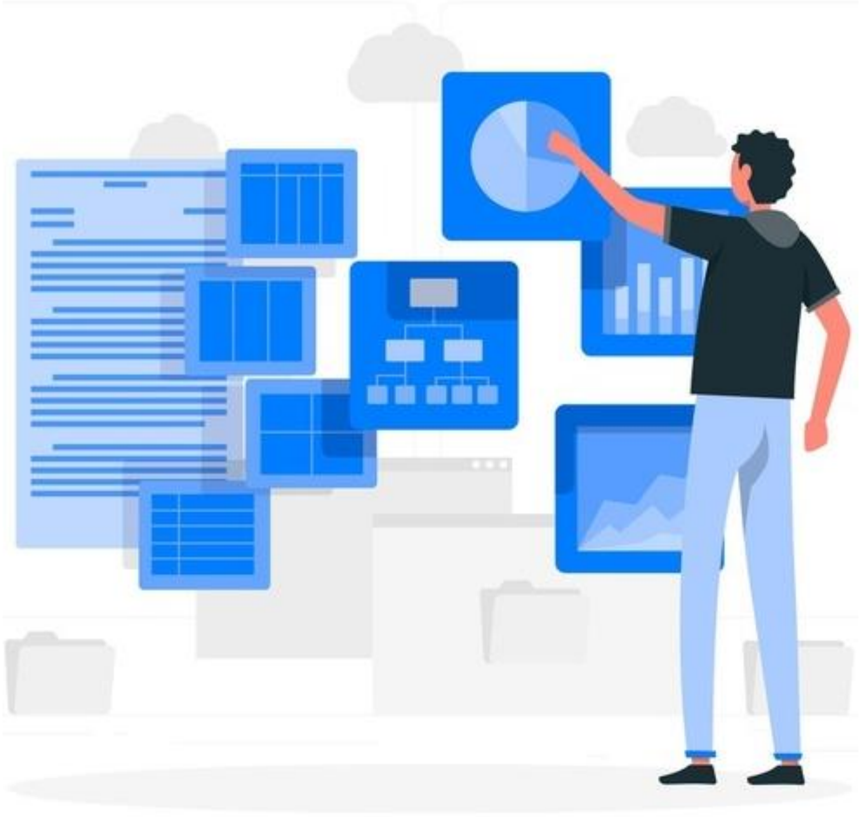
Boruta is a feature selection algorithm that works as a wrapper around random forest.



Predictive modeling is crucial to get the most important feature, also known as feature selection.

Boruta

Boruta becomes useful to select features, when the data set contains many features.



It eliminates features in each iteration while fitting the random forest model.

It minimizes the error and results in over-pruned input data that throws away some relevant features.

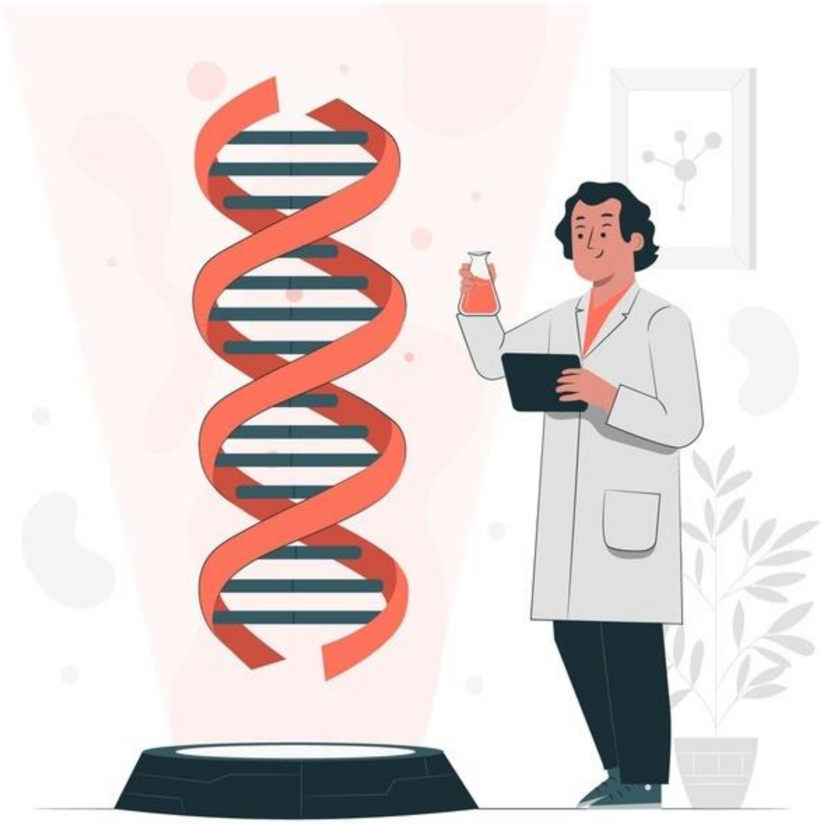
Boruta

Boruta analyzes all the high and weak features connected to the decision variable.

It captures all the features relevant to the dependent or target variable by following a certain feature selection method.

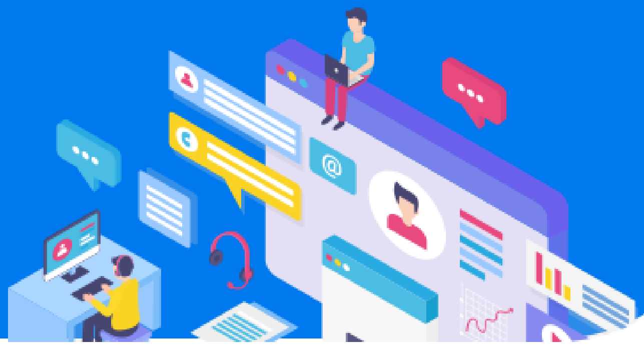
Applications of Boruta

Boruta is used majorly in the biomedical industry.



It helps the doctors find the genes that are causing health-related problems, which is otherwise complicated.

Assisted Practices



Let’s understand the topic below using Jupyter Notebook.

- 5.16_Automatic Feature Selection with Boruta

Note: Please download the pdf files for each topic mentioned above from the Reference Material section.

Support Vector Machine

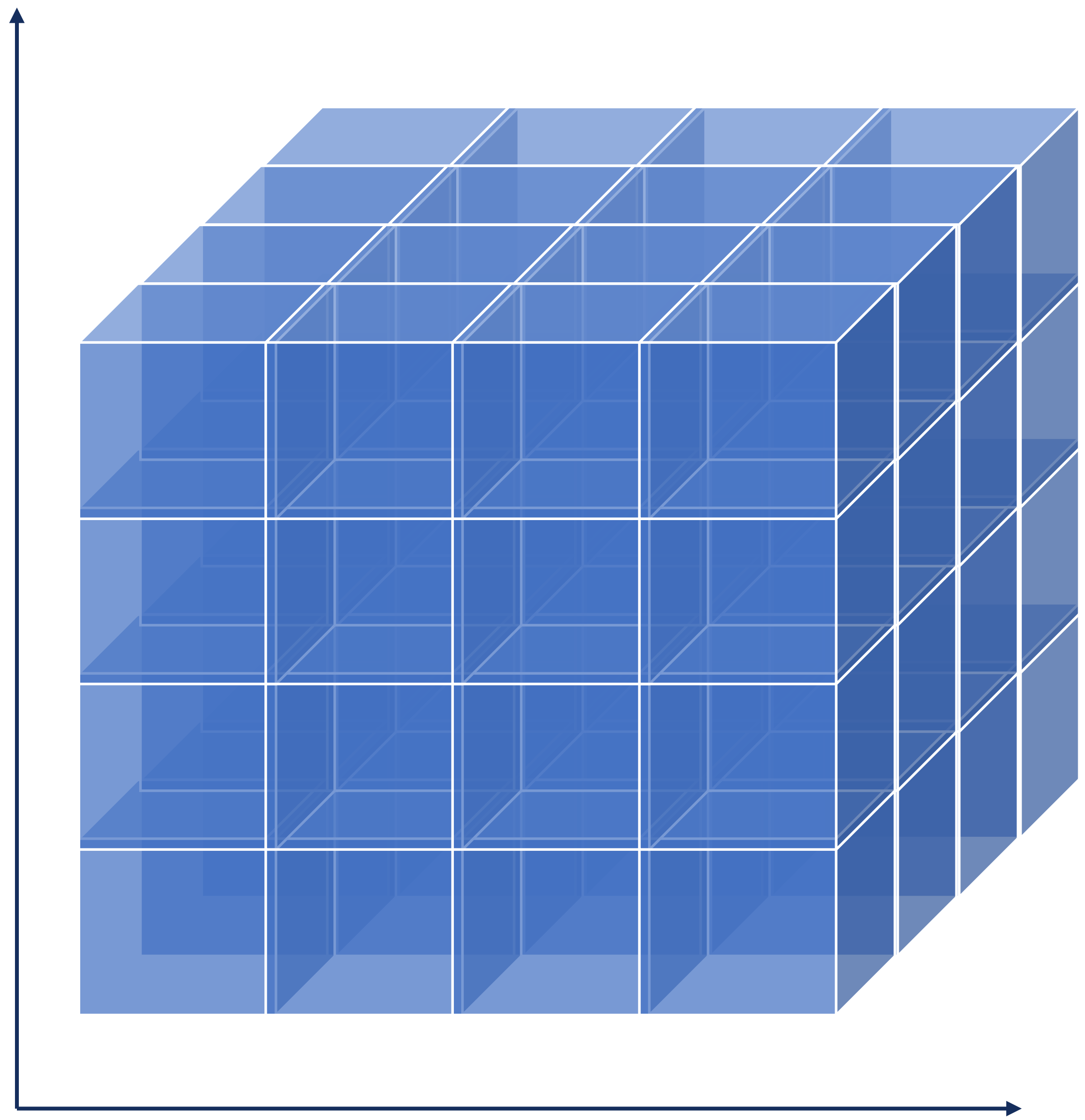
Support Vector Machine (SVM)

It is a supervised machine learning algorithm used for both classification and regression challenges.

It is mainly utilized in classification-related problems.

Support Vector Machine (SVM)

Let’s plot the data points in n-dimensional space, where n is the number of features in data.

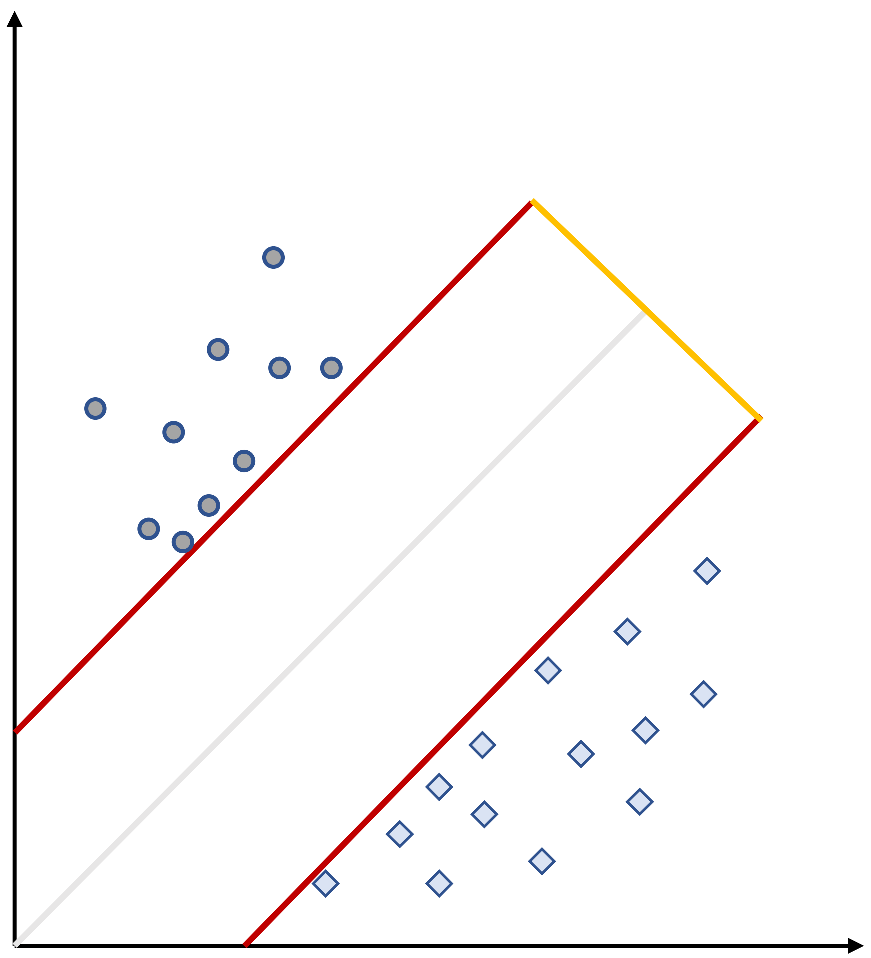


Perform classification using hyperplane

Hyperplane helps differentiate the classes

Support Vector Machine (SVM)

Following is a support vector algorithm:



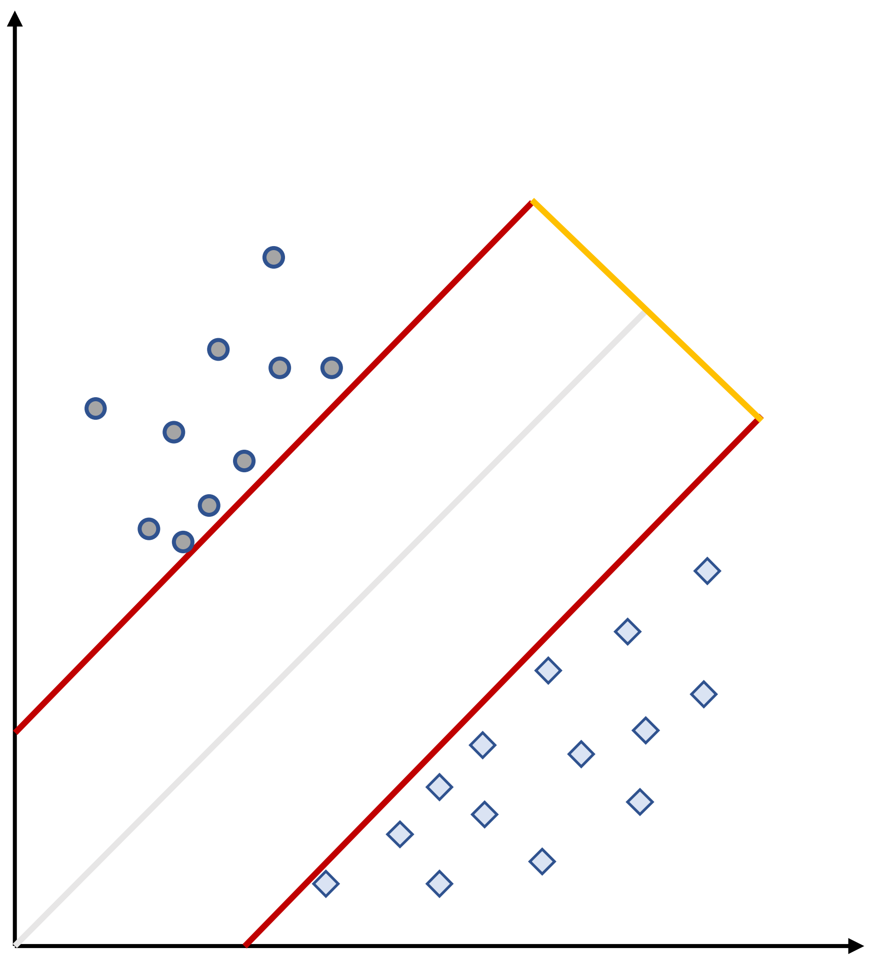
Legend

- Support Vector
- Max. Margin Hyperplane
- Max. Margin
- Target = No
- Target = Yes

It provides analysis of data for classification and regression.

Support Vector Machine (SVM)

In this, plotting is done in the n^{th} dimension.



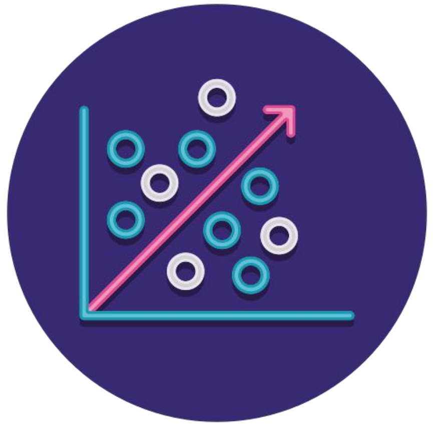
The value of the data point and the value of the coordinate in the dimension are the same.

The hyperplane helps to classify the two classes of data.

In KNN, a kernel function is used to map lower-dimensional data to higher-dimensional data.

Types of Support Vector Machine

There are two types of support vector machines:



Linear SVM: separates the data in a linear format

If the data set is separated into two using a straight line, data is linearly separable.

Nonlinear SVM: used when data is nonlinearly separated

If the data set cannot be separated into two using a straight line, data is nonlinear.

Applications of SVM

It is used to classify different avatars that players use in games.
Example: to classify the positive avatars and the negative avatars in a game

It is best used to recommend ads.
Example: only showing ads relevant to kids on kid’s channels

It is used to classify cataracts, a medical condition during which the lens of the patient becomes progressively opaque, leading to blurred vision.

Assisted Practices

Let's understand the topic below using Jupyter Notebook.

- 5.18_Applying Support Vector Machine

Note: Please download the pdf files for each topic mentioned above from the Reference Material section.

ASSISTED PRACTICE

Cohen's Kappa

Cohen's Kappa

Cohen's kappa is a statistical model used to:

Measure the reliability of two raters rating the same item

Identify how frequently the raters are in agreement

Compare the predictions done by the machine learning model

Cohen's Kappa

Software packages and libraries that provide Cohen’s kappa are:

- Caret
- Weka
- Scikit-learn

Cohen's Kappa

Cohen's kappa is calculated based on the confusion matrix but is complex as it takes an imbalance in class distribution to calculate accuracy.

$$k = \frac{p_0 - p_e}{1 - p_e}$$

where

P_0 - Accuracy of the model

P_e - Measure of the agreement between the model predictions and the actual class values

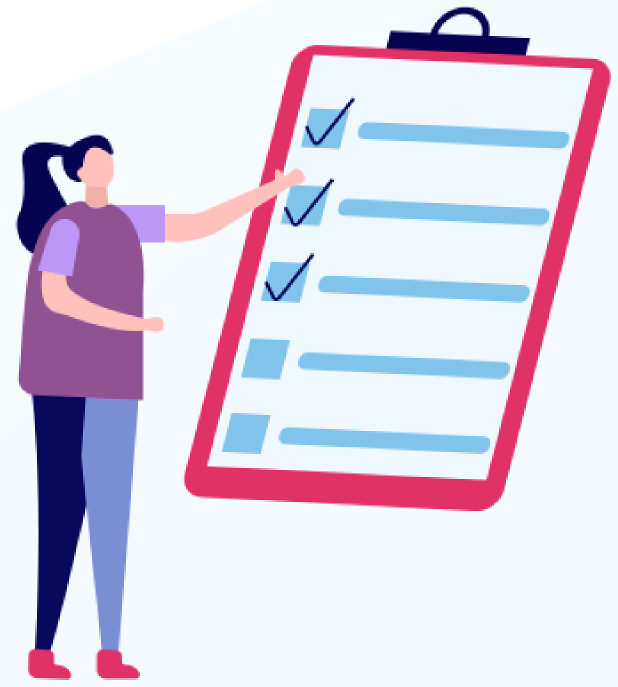
Cohen's Kappa: Key Points

Some key points about Cohen's kappa are:

- It is an excellent measure that handles multi-class and imbalanced class problems.
- It decreases with an increase in the difference between predicted and actual target classes.
- Balanced data gives a higher Cohen's kappa value.
- It is more informative than overall accuracy when working with unbalanced data.

Key Takeaways

- Classification algorithms are used to categorize data into specific classes or categories based on a set of prior observations.
- Performance metrics such as accuracy, precision, recall and specificity are used to evaluate the effectiveness of classification models.
- Naive Bayes classifier is a machine learning model that segregates different objects on the basis of certain features of variables.
- The gradient descent algorithm is an iterative process that works by descending a function's slope until it reaches the minimum.
- K-nearest neighbors (KNN) and decision trees are the two main machine learning algorithms that support supervised learning.



Key Takeaways

- Random forest is used to solve regression and classification problems that combines multiple classifiers to solve a complex problem.
- Boruta is a feature selection algorithm that works as a wrapper around random forest and is useful for selecting features when the data set contains many of them.
- Support vector machine (SVM) is a supervised machine learning algorithm used for classification and regression challenges.
- Cohen's kappa is used to compare the predictions done by the machine learning model.

Knowledge Check

Knowledge Check

1

What are the types of classification algorithms?

A. Binary classification, multi-class classification, multi-label classification and imbalance classification

B. Linear regression, logistic regression, decision trees and k-nearest neighbors

C. Neural networks, random forests and support vector machines

D. Naive Bayes, multinomial Bayes and Bernoulli Bayes

Knowledge Check 1

What are the types of classification algorithms?

A. Binary classification, multi-class classification, multi-label classification and imbalance classification

B. Linear regression, logistic regression, decision trees and k-nearest neighbors

C. Neural networks, random forests and support vector machines

D. Naive Bayes, multinomial Bayes and Bernoulli Bayes

The correct answer is **A**

The four types of classification algorithms are binary classification, multi-class classification, multi-label classification and imbalance classification.

Knowledge Check

2

What is the advantage of stochastic gradient descent over gradient descent algorithm?

A. It guarantees to reach the minimum point of the function.

B. It reduces the number of computations required to complete the algorithm.

C. It provides more accurate results than gradient descent.

D. It is less computationally expensive than gradient descent.

Knowledge Check

2

What is the advantage of stochastic gradient descent over gradient descent algorithm?

A. It guarantees to reach the minimum point of the function.

B. It reduces the number of computations required to complete the algorithm.

C. It provides more accurate results than gradient descent.

D. It is less computationally expensive than gradient descent.

The correct answer is **B**

It reduces the number of computations required to complete the algorithm.

Knowledge Check

3

What is the significance of Boruta in feature selection?

A. It recursively eliminates features in each iteration while fitting the random forest model

B. It analyzes all the high and weakly connected features to the decision variable

C. It captures all the features relevant to the dependent or target variable

D. All of the above

Knowledge Check

3

What is the significance of Boruta in feature selection?

A. It recursively eliminates features in each iteration while fitting the random forest model

B. It analyzes all the high and weakly connected features to the decision variable

C. It captures all the features relevant to the dependent or target variable

D. All of the above

The correct answer is **D**

Boruta is valuable for feature selection in data sets with many features as it recursively eliminates the irrelevant ones.

Thank You!