JOSÉ RIZAL UNIVERSITY

Find your Inner Hero

husay rizal

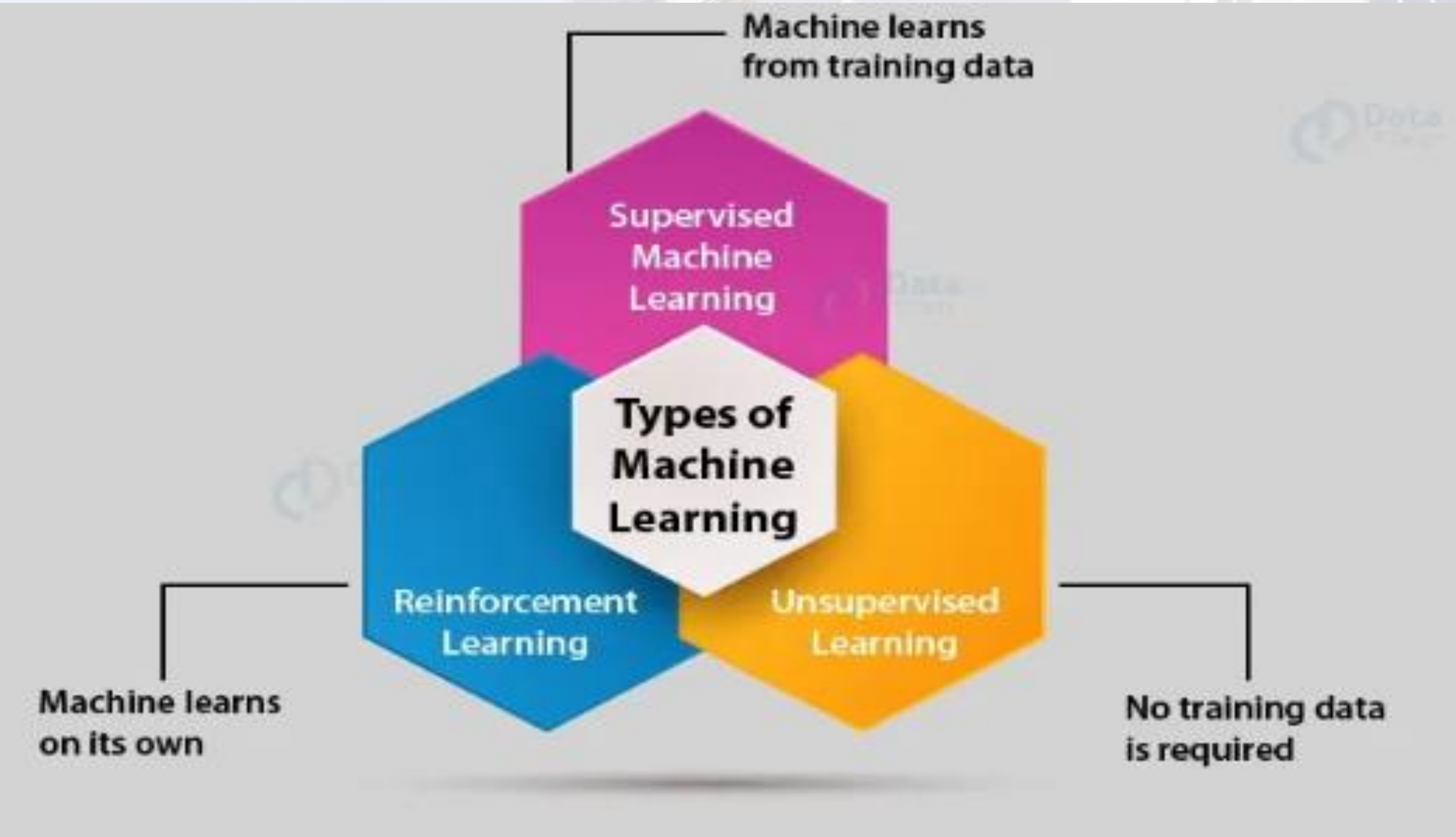
ITC C307– Foundations of Business Analytics

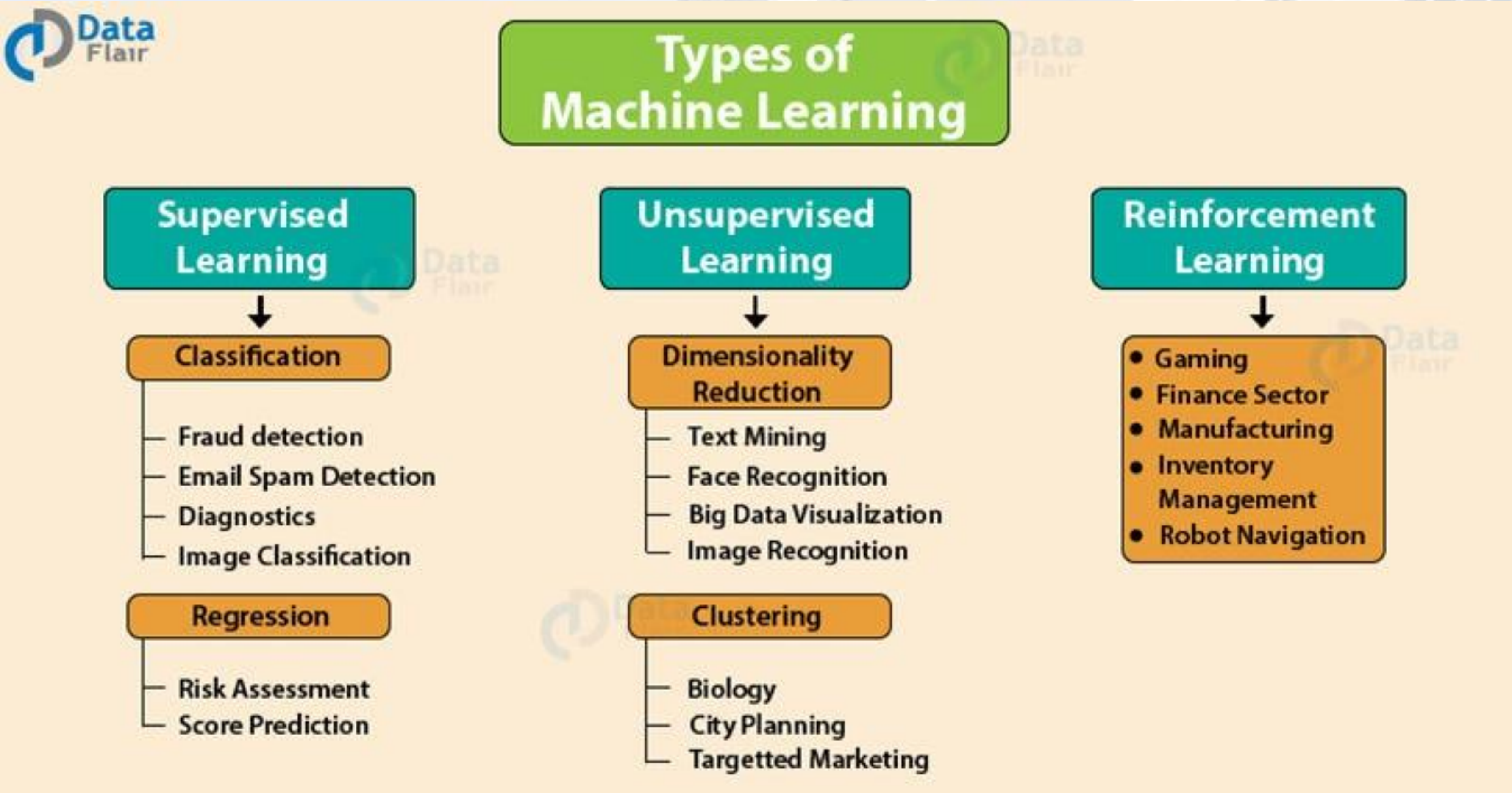
Supervised Learning: Classification

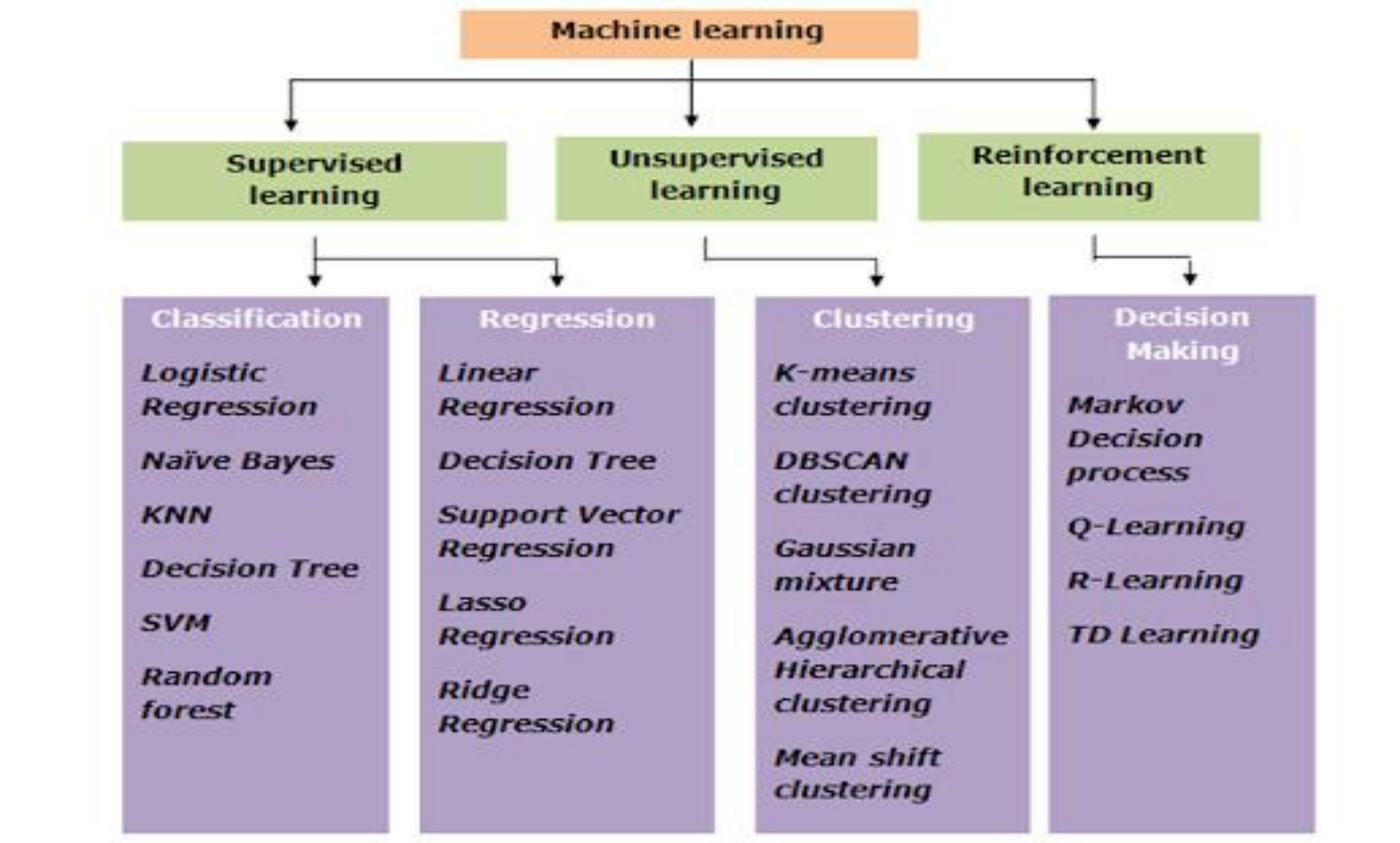
Prepared by:
Mengvi P. Gatpandan, DIT

Learning Objectives

- We are learning to.....
  - Classification definition
  - Classification task
  - Classification algorithms
  - Accuracy model


Machine learns
from training data
Supervised
Machine
Learning
Types of
Machine
Learning
Reinforcement
Learning
Unsupervised
Learning
Machine learns
on its own
No training data
is required


Data Flair
Types of Machine Learning
Supervised Learning
Classification
Fraud detection
Email Spam Detection
Diagnostics
Image Classification
Regression
Risk Assessment
Score Prediction
Unsupervised Learning
Dimensionality Reduction
Text Mining
Face Recognition
Big Data Visualization
Image Recognition
Clustering
Biology
City Planning
Targetted Marketing
Reinforcement Learning
• Gaming
• Finance Sector
• Manufacturing
• Inventory Management
• Robot Navigation


Machine learning
Supervised learning
Unsupervised learning
Reinforcement learning
Classification
Logistic Regression
Naïve Bayes
KNN
Decision Tree
SVM
Random forest
Regression
Linear Regression
Decision Tree
Support Vector Regression
Lasso Regression
Ridge Regression
Clustering
K-means clustering
DBSCAN clustering
Gaussian mixture
Agglomerative Hierarchical clustering
Mean shift clustering
Decision Making
Markov Decision process
Q-Learning
R-Learning
TD Learning

Classification

- Predicting the value of a categorical variable (target or class)
- This is done by building a model based on one or more numerical and/or categorical variables (predictors, attributes or features)
- An observations can be described by a fixed set of quantifiable properties called Explanatory variables or features.
  - Example, a doctor visits could result in the following features: weight, gender, age, cell count, mental state (bad, neutral, good, great) blood pressure, and so on.
  - Text documents will have a set of features that defines the number of occurrences of each Word or n-gram in the corpus documents.

Classifier

- A machine learning algorithm or mathematical function that maps input data to a category is known as a ***Classifier.***
- Most algorithms are best applied to Binary Classification.

Types of Classification Task – Binary Classification

- Those classification jobs with only two class labels are referred to as binary classification.
  - Examples: Prediction of conversion (buy or not), Churn forecast (churn or not), Detection of spam email (spam or not), and so on.
  - Binary classification problems often require two classes, one representing the normal state and the other representing the aberrant state.
  - A model that forecasts a Bernoulli probability distribution for each case is frequently used to represent a binary classification task.
  - The discrete probability distribution known as the *Bernoulli distribution* deals with the situation where an event has a binary result of either 0 or 1. In terms of classification, this indicates that the model forecasts the likelihood that an example would fall within class 1, or the abnormal state.
  - ***The following are well-known binary classification algorithms: Logistic Regression, Decision Tress, Simple Bayes, and Support Vector Machines***

Types of Classification Task – Multi Class Classification

- Multi-class labels are used in classification tasks referred to as multi-class classification.
  - Examples: Categorization of faces, Classifying plant species, Character recognition using optical.
  - The multi-class classification does not have the idea of normal and abnormal outcomes, in contrast to binary classification. Instead, instances are grouped into one of several well-known classes.
  - In some cases, the number of class labels could be rather high. In a facial recognition system, for instance, a model might predict that a shot belongs to one of thousands or tens of thousands of faces.
  - Multiclass classification tasks are frequently modeled using a model that forecasts a Multinoulli probability distribution for each example.
  - An event that has a categorical outcome, such as K in 1, 2, 3,..., K, is covered by the Multinoulli distribution, which is a discrete probability distribution. In terms of classification, this implies that the model forecasts the likelihood that a given example will belong to a certain class label.

Types of Classification Task – Multi Class Classification

- For multi-class classification, many binary classification techniques are applicable.
- The following well-known algorithms can be used for multi-class classification:
- Progressive Boosting
- Choice trees
- Nearest K Neighbors
- Rough Forest
- Simple Bayes

Types of Classification Task – Multi Label Classification

- Multi-label classification problems are those that feature two or more class labels and allow for the prediction of one or more class labels for each example.
  - Think about the photo classification example. Here a model can predict the existence of many known things in a photo, such as “person”, “apple”, "bicycle," etc. A particular photo may have multiple objects in the scene.
  - This greatly contrasts with multi-class classification and binary classification, which anticipate a single class label for each occurrence.
  - Multi-label classification problems are frequently modeled using a model that forecasts many outcomes, with each outcome being forecast as a Bernoulli probability distribution. In essence, this approach predicts several binary classifications for each example.
  - It is not possible to directly apply multi-label classification methods used for multi-class or binary classification. *The so-called multi-label versions of the algorithms, which are specialized versions of the conventional classification algorithms, include: Multi-label Gradient Boosting, Multi-label Random Forests, and Multi-label Decision Trees*

Types of Classification Task – Imbalanced-Classification

- The term "imbalanced classification" describes classification jobs where the distribution of examples within each class is not equal.
- A majority of the training dataset's instances belong to the normal class, while a minority belong to the abnormal class, making imbalanced classification tasks binary classification tasks in general.
- Although they could need unique methods, these issues are modeled as binary classification jobs.
- By oversampling the minority class or under sampling the majority class, specialized strategies can be employed to alter the sample composition in the training dataset.
- Examples: SMOTE Oversampling and Random Under sampling

Learners in Classification Problems

- **Lazy Learners**
  - It first stores the training dataset before waiting for the test dataset to arrive. When using a lazy learner, the classification is carried out using the training dataset's most appropriate data. Less time is spent on training, but more time is spent on predictions. Some of the examples are case-based reasoning and the KNN algorithm.
- **Eager Learners**
  - Before obtaining a test dataset, eager learners build a classification model using a training dataset. They spend more time studying and less time predicting. Some of the examples are ANN, naive Bayes, and Decision trees.

Classification Data

- Training
  - It has a set of variables that need to set (trained). Different classifiers have different algorithms to optimize this process.
- Overfitting
  - Danger! The model fits only the data in was trained on. New data is completely foreign.
- Split the data into In-Sample (training) and Out-Of-Sample (test)

Classification-Based Algorithms

- Covariance Matrix
  - Linear Discriminant Analysis
  - Logistic Regression
- Frequency Table
  - Decision Tree
  - Naïve Bayesian
- Similarity Functions
  - K Nearest Neighbors
- Others
  - Artificial Neural Network
  - Support Vector Machine

Classification Types - Logistic Regression

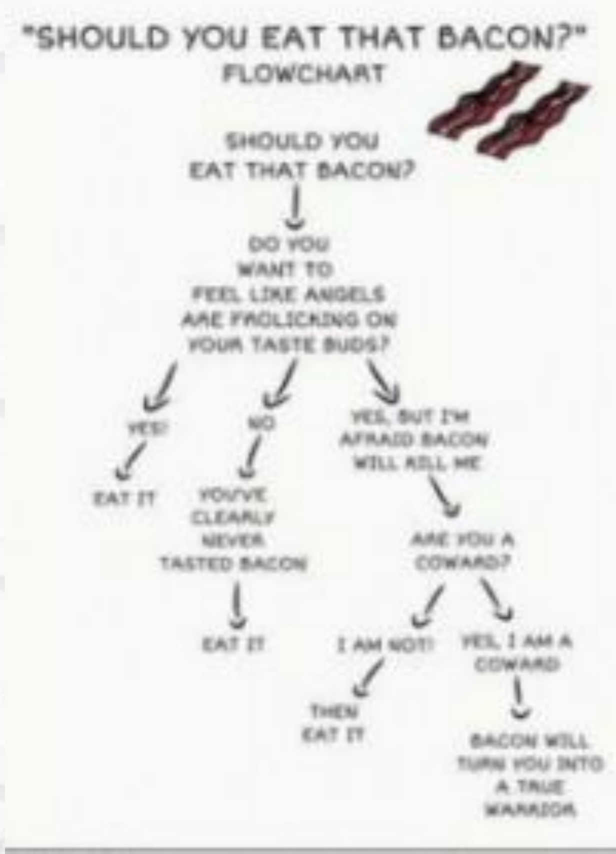
- It is a supervised learning classification technique that forecasts the likelihood of a target variable. There will only be a choice between two classes.
- Data can be coded as either one or yes, representing success, or as 0 or no, representing failure. The dependent variable can be predicted most effectively using logistic regression.
- When the forecast is categorical, such as true or false, yes or no, or a 0 or 1, you can use it.
- A logistic regression technique can be used to determine whether or not an email is spam.

Classification Types – Decision Tree

- Builds classification or regression models in the form of tree structure.
- It breaks down a data set into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed.
- The final result is a tree with decision nodes and leaf nodes.
  - A **decision no**de has two or more branches
  - **Lead node** represents a classification or decision.
- The topmost decision node is tree which corresponds to the best predictor called **root node**.

Classification Types – Decision Tree

- Decision trees can handle both categorical and numerical data.
- Can be learned by consecutively splitting the data on an attribute pair using Recursive Partitioning.
- Can generate multiple decision tree to improve accuracy (Random Forest)



Classification Types – Random Forest

- The random forest algorithm is an extension of the Decision Tree algorithm where you first create a number of decision trees using training data and then fit your new data into one of the created ‘trees’ as a ‘random forest’. It averages the data to connect it to the nearest tree data based on the data scale. These models greatly improve the decision tree’s problem of unnecessarily forcing data points within a category.

Classification Types – Random Forest

- **Are decision trees in Random Forest different from regular decision trees?**
- For example, if you wanted to predict how much a bank's customers will use a specific service a bank provides with a single decision tree, you would gather up how often they've used the bank in the past and what service they utilized during their visits. You would add some features that describe that customer's decisions. The decision tree will generate rules to help predict whether the customer will use the bank's service.
- If you inputted that same dataset into a **Random Forest**, the algorithm would build multiple trees out of randomly selected customer visits and service usage. Then it would output the average results of each of those trees.

Classification Types – Naïve Bayesian

- Works based on Bayes' Theorem
- Why is it called Naïve? Because it assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature.
- Easy to build
- Useful for vast data sets.

Classification Types – Naïve Bayesian

- It determines whether a data point falls into a particular category. It can be used to classify phrases or words in text analysis as either falling within a predetermined classification or not.

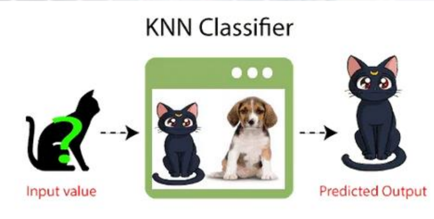
Text	Tag
"A great game"	Sports
"The election is over"	Not Sports
"What a great score"	Sports
"A clean and unforgettable game"	Sports
"The spelling bee winner was a surprise"	Not Sports

Classification Types – K Nearest Neighbors

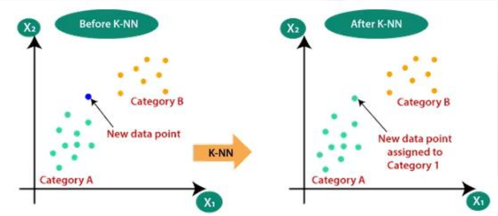
- It calculates the likelihood that a data point will join the groups based on which group the data points closest to it are a part of.
- When using k-NN for classification, you determine how to classify the data according to its nearest neighbor. It tries to predict the correct class for the test data by calculating the distance between the test data and all the training points. Then select the K number of points closest to the test data.
- The KNN algorithm calculates the probability of the test data belonging to the classes of ‘K’ training data, and class holds the highest probability will be selected.
- In the regression case, the value is the mean of the ‘K’ selected training points.

Classification Types – K Nearest Neighbors

- Suppose, we have an image of a creature that looks similar to cat and dog, but we want to know either it is a cat or dog. So for this identification, we can use the KNN algorithm, as it works on a similarity measure. Our KNN model will find the similar features of the new data set to the cats and dogs images and based on the most similar features it will put it in either cat or dog category.



- Suppose there are two categories, i.e., Category A and Category B, and we have a new data point x1, so this data point will lie in which of these categories. To solve this type of problem, we need a K-NN algorithm. With the help of K-NN, we can easily identify the category or class of a particular dataset. Consider the below diagram:



Classification Types – Support Vector Machine

- It is a popular supervised machine-learning technique for classification and regression problems. It goes beyond X/Y prediction by using algorithms to classify and train the data according to polarity. You can use specific types of SVMs for particular machine learning problems, like support vector regression (SVR), which is an extension of support vector classification (SVC).
- SVMs are different from other classification algorithms because of the way they choose the decision boundary that maximizes the distance from the nearest data points of all the classes. The decision boundary SVMs create is the maximum margin classifier or the maximum margin hyperplane.

Classification Types – Support Vector Machine

- **Types of SVMs**
- There are two different types of SVMs, each used for different things:
- **Simple SVM:** Typically used for linear regression and classification problems.
- **Kernel SVM:** It has more flexibility for non-linear data because you can add more features to fit a hyperplane instead of a two-dimensional space.

JOSÉ RIZAL UNIVERSITY
Find your Inner Hero

husay rizal

How do we evaluate classifier performance?

- Accuracy is used in classification problems to tell the percentage of correct predictions made by a model.
- Accuracy score in machine learning is an evaluation metric that measures the number of correct predictions made by a model in relation to the total number of predictions made. We calculate it by dividing the number of correct predictions by the total number of predictions.

$$Accuracy = \frac{Number\ of\ Correct\ Predictions}{Total\ Number\ of\ Predictions}$$

How do we evaluate classifier performance?

- Precision: Percentage of correct predictions of a class among all predictions for that class.
- Recall: Proportion of correct predictions of a class and the total number of occurrences of that class.
- F-score: A single metric combination of precision and recall.
- Confusion matrix: A tabular summary of True/False Positive/Negative prediction rates.
- ROC curve: A binary classification diagnostic plot.

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

		Predicted	
		Negative (N) -	Positive (P) +
Actual	Negative -	True Negative (TN)	False Positive (FP) Type I Error
	Positive +	False Negative (FN) Type II Error	True Positive (TP)