# Machine Learning

## What is Machine Learning

The use and development of computer systems that can learn and adapt without following explicit instructions, by using algorithms and statistical models to analyze and draw inferences from patterns in data.

## When we use Machine Learning

* When you have a problem that requires many long lists of rules to find the solution. In this case, machine-learning techniques can simplify your code and improve performance.
* Very complex problems for which there is no solution with a traditional approach.
* Non- stable environments’: machine-learning software can adapt to new data.

## **Data Types from A Machine Learning Perspective**

### **NUMERICAL DATA**

* Numerical data is any data where data points are exact numbers. Statisticians also might call numerical data, quantitative data.
* Numerical data can be characterized by continuous or discrete data. Continuous data can assume any value within a range whereas discrete data has distinct values.

Diagram

Description automatically generated

### **CATEGORICAL DATA**

* Categorical data represents characteristics, such as a hockey player’s position, team, hometown. Categorical data can take numerical values. For example, maybe we would use 1 for the colour red and 2 for blue. But these numbers don’t have a mathematical meaning. That is, we can’t add them together or take the average.
* In the context of super classification, categorical data would be the class label. This would also be something like if a person is a man or woman, or property is residential or commercial.

### **TIME-SERIES DATA**

* Time series data is a sequence of numbers collected at regular intervals over some period of time. It is very important, especially in particular fields like finance. Time series data has a temporal value attached to it, so this would be something like a date or a timestamp that you can look for trends in time.

Chart, line chart

Description automatically generated

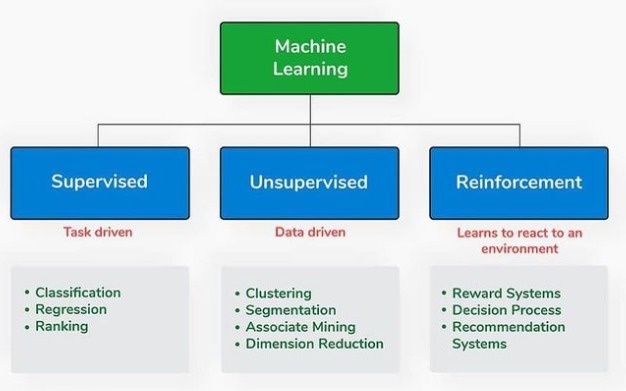
### **TEXT**

* Text data is basically just words. A lot of the time the first thing that you do with text is you turn it into numbers using some interesting functions like the bag of words formulation.

A picture containing text, accessory

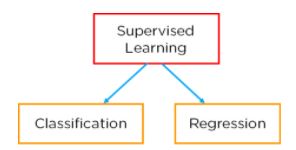
Description automatically generated

## Types of Systems of Machine Learning



## Supervised & Unsupervised Learning

### Supervised Learning



In this type of machine-learning system, the data that you feed into the algorithm, with the desired solution, are referred to as “labels.”

#### Regression

##### Regression Model

– relationship between dependent and independent variable and the output is continuous. (Predict a **number**)

* Linear regression-
* Decision Tree-
* Random Forest-
* Neural Network-

#### Classification

Classification is defined as the process of **recognition**, understanding, and grouping of objects and ideas into preset categories a.k.a “sub-populations.” With the help of these pre-categorized training datasets, classification in machine learning programs leverage a wide range of algorithms to classify future datasets into respective and relevant categories.

Classification algorithms used in machine learning utilize input training data for the purpose of predicting the likelihood or probability that the data that follows will fall into one of the predetermined categories. One of the most common applications of classification is for filtering emails into “spam” or “non-spam”, as used by today’s top email service providers.

In short, classification is a form of “pattern recognition,”. Here, classification algorithms applied to the training data find the same pattern (similar number sequences, words or sentiments, and the like) in future data sets.

We will explore classification algorithms in detail, and discover how a text analysis software can perform actions like sentiment analysis - used for categorizing unstructured text by opinion polarity (positive, negative, neutral, and the like).

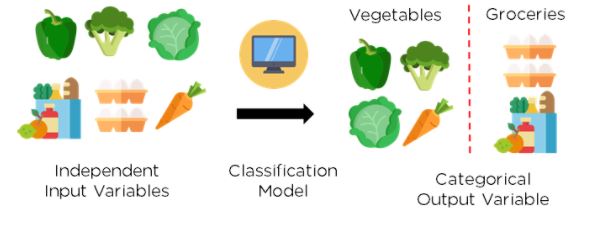


Figure 2: Classification of vegetables and groceries

##### What is Classification Model/Algorithm?

Based on training data, the Classification algorithm is a Supervised Learning technique used to categorize new observations. In classification, a program uses the dataset or observations provided to learn how to categorize new observations into various classes or groups. For instance, 0 or 1, red or blue, yes or no, spam or not spam, etc. Targets, labels, or categories can all be used to describe classes. The Classification algorithm uses labeled input data because it is a supervised learning technique and comprises input and output information. A discrete output function (y) is transferred to an input variable in the classification process (x).

In simple words, classification is a type of pattern recognition in which classification algorithms are performed on training data to discover the same pattern in new data sets.

Classification defined as the process of recognition, understanding, and grouping of objects and ideas into preset categories a.k.a “sub-populations.” With the help of these pre-categorized training datasets, classification in machine learning programs leverage a wide range of algorithms to classify future datasets into respective and relevant categories.

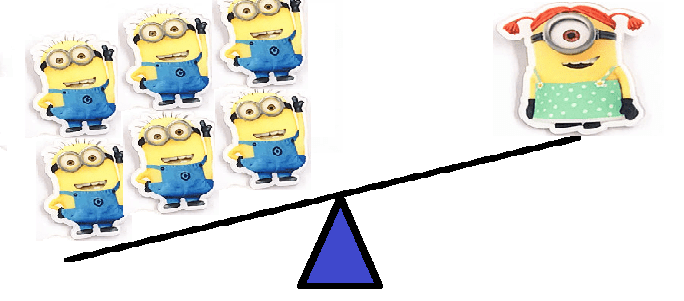
Classification Model/algorithm is a Supervised Learning technique used to categorize new observations. In classification, a program uses the dataset or observations provided to learn how to categorize new observations into various classes or groups. For instance, 0 or 1, red or blue, yes or no, spam or not spam, etc.

– the output is discrete. Predict a category

* + Logistics Regression
  + Support Vector Machine (SVM)
  + Naïve Bayes
  + Decision Tree
  + Random Forest
  + Neural Network
  + K Nearest Neighbors

###### **Imbalanced Classification**

Application

Description automatically generated

The term "imbalanced classification" describes classification jobs where the distribution of examples within each class is not equal.

