# PRIOR KNOWLEDGE

| Date | 15 November 2022 |
|---|---|
| Team ID | PNT2022TMID21644 |
| Project Name | Crude Oil Price Prediction |

**One should have knowledge on the following Concepts:**

**Supervised and unsupervised learning:**

- Supervised learning has the presence of a supervisor as a teacher. Basically, supervised learning is when we teach or train the machine using data that is well labelled. Which means some data is already tagged with the correct answer. After that, the machine is provided with a new set of examples(data) so that the supervised learning algorithm analyses the training data (set of training examples) and produces a correct outcome from labelled data.

  Supervised learning is classified into two categories of algorithms:

  - **Classification**: A classification problem is when the output variable is a category, such as "Red" or "blue", "disease" or "no disease".
  - **Regression**: A regression problem is when the output variable is a real value, such as "dollars" or "weight"

- Unsupervised learning is the training of a machine using information that is neither classified nor labeled and allowing the algorithm to act on that information without guidance. Here the task of the machine is to group unsorted information according to similarities, patterns, and differences without any prior training of data.

  Unlike supervised learning, no teacher is provided that means no training will be given to the machine. Therefore, the machine is restricted to find the hidden structure in unlabeled data by itself.

**Regression Classification and Clustering:**

- Clustering is an unsupervised technique. With clustering, the algorithm tries to find a pattern in data sets without labels associated with it. This could be a clustering of buying behavior of customers. Features for this would be the household income, age, … and clusters of different consumers could then be built.

- In contrast to clustering, classification is a supervised technique. Classification algorithms look at existing data and predicts what a new data belongs to. Classification is used for spam for years now and these algorithms are more or less mature in classifying something as spam or not. With machine data, it could be used to predict a material quality by several known parameters (e.g., humidity, strength, color, …). The output of the material prediction would then be the quality type (either "good" or "bad" or a number in a defined space like 1-10). Another well-known sample is if someone would survive the titanic – classification is done by "true" or "false" and input parameters are "age", "sex", "class". If you would be 55, male and in 3rd class, chances are low, but if you are 12, female and in first class, chances are rather high.

- Regression is often confused with clustering, but it is still different from it. With a regression, no classified labels (such as good or bad, spam or not spam, …) are predicted. Instead, regression outputs continuous, often unbound, numbers. This makes it useful for financial predictions and alike. A commonly known sample is the prediction of housing prices, where several values (FEATURES!) are known, such as distance to specific landmarks, plot size. The algorithms could then predict a price for your house and the amount you can sell it for.

**Artificial Neural Networks:**

- Artificial neural networks (ANNs) are comprised of a node layers, containing an input layer, one or more hidden layers, and an output layer. Each node, or artificial neuron, connects to another and has an associated weight and threshold. If the output of any

individual node is above the specified threshold value, that node is activated, sending data to the next layer of the network. Otherwise, no data is passed along to the next layer of the network.

- Neural networks rely on training data to learn and improve their accuracy over time. However, once these learning algorithms are fine-tuned for accuracy, they are powerful tools in computer science and artificial intelligence, allowing us to classify and cluster data at a high velocity. Tasks in speech recognition or image recognition can take minutes versus hours when compared to the manual identification by human experts. One of the most well-known neural networks is Google's search algorithm.

**Convolution Neural Networks:**

- Convolutional neural networks are distinguished from other neural networks by their superior performance with image, speech, or audio signal inputs. They have three main types of layers, which are:

  - Convolutional layer
  - Pooling layer
  - Fully connected (FC) layer

- The convolutional layer is the first layer of a convolutional network. While convolutional layers can be followed by additional convolutional layers or pooling layers, the fully connected layer is the final layer. With each layer, the CNN increases in its complexity, identifying greater portions of the image. Earlier layers focus on simple features, such as colors and edges. As the image data progresses through the layers of the CNN, it starts to recognize larger elements or shapes of the object until it finally identifies the intended object.