**Predictive Analytics:** Technology that learns from data to predict the future behaviour of individuals in order to drive better decisions.

**Predictive Model:** Used to analyse data and historical information to make predictions about the future. It takes characteristics (e.g. age, occupation etc) as input and returns a prediction (e.g. If person will purchase item).

**Supervised Machine Learning and Training:** Helps the system understand what the essential characteristics are by providing a set of inputs and outputs. This allows the algorithm to predict the outputs for future inputs. These new input and outputs are then fed back into the algorithm. Training produces a machine learning model that can then be used to make predictions.

**Feature/Characteristic**: Features are individual independent variables that act as the input in your system.

**Target**: The target is the output of the input variables.

**Label**: Labels are the final output. You can also consider the output classes to be the labels. When data scientists speak of labelled data, they mean groups of samples that have been tagged to one or more labels.

**Ill Posed Problem:** For most scenarios there are many models. A single consistent model cannot be found based on the sample training data alone, therefore machine learning is an ill posed problem.

**Inductive Bias:** The set of assumptions that the model uses to predict outputs given inputs that it has not encountered. The model that generalises beyond the dataset the best must be found.

**Restriction Bias:** Constrains the set of models that the algorithm will consider during training.

**Preference Bias:** Guides the algorithm to prefer certain models over others.

**Underfitting:** When the model selected by the algorithm is too simplistic to represent the relationship the characteristics and the target feature.

**Overfitting:** When the model selected by the algorithm is too complex so that it is sensitive to noise in the data.

**Regularization:** Regularisation is the method to estimate a preferred complexity of the machine learning model so that the model generalises, and the over-fit/under-fit problem is avoided. This is done by adding a penalty on the different parameters of the model thereby reducing the freedom of the model.

**CRISP-DM:** Cross-industry standard process for data mining.

* **Business Understanding Phase:** An In-depth analysis of the business objectives and needs. A plan to proceed is set up.
* **Data Understanding Phase:** Data is collected from all available sources. The properties of the data are examined. The completeness and accuracy of the data is examined.
* **Preparation of Data Phase:** Data is organised into specific structures known as Analytics Base Tables (ABT). All data is converted into a well formed ABT which machine learning models can use.
* **Modelling Phase:** Different machine learning algorithms are used to build a range of prediction models. The best model is selected.
* **Evaluation Phase:** The model is tested to see if it suffers from under fitting or overfitting and that it can make accurate predictions.
* **Deployment Phase:** The machine learning model is integrated into an organisations process.

**Algorithm Examples:**

* **Probability Based Learning:**
* **Similarity Based Learning:**
* **Information Based Learning:**
* **Error Based Learning:**