Glossary

Supervised learning-
When a model/machine is trained with input and output data, until the model/machine can detect patterns.

Classification-
Type of supervised learning task, where input data is associated with a distinct output variable, such as true or false.

Decision Tree-
A supervised learning task that can be used to for both classification and regression tasks. It uses a tree like model that branch out to 'leaf nodes' where it will make a final decision.

Random Forest-
An 'Ensemble learning' method that is supervised machine learning algorithm. It is a collection of 'Decision Trees' which are averaged on different parts of a subset of a data/dataset to come to a final decision.

Target variable-
This is the variable in which the machine learning will be compared to, the machine learning algorithm will aim to predict identical results to this variable.

Binary-
A number system with only two possible outcomes, often denoted as '1' or '0' or 'True'. 'False'

Standardisation-
The Process of Scaling variables to have a mean of 0 and a standard deviation of 1.

Ensemble Learning-
Machine learning technique, which combines multiple models (i.e. Random Forest).

Root Node-
In a Decision tree, and referrers to the starting/entire dataset.

Internal Nodes-
Points within a Decision Tree, which split the data into subsets depending on conditions.

Leaf Nodes-
Also known as Terminal nodes, where the final decision/classification is made.

Pearsons Correlation-
A statistical measure of relationship between variables

Overfitting-
A Machine learning behaviour, that gives accurate predictions for training data but not for unseen data.