|  |
| --- |
| Photo displaying partial image of two pie charts on a canvas-textured page |
| Machine Learning Algorithms |
| |  |  |  | | --- | --- | --- | | Fiona Warigon |  | Data Analysis | |

# **Machine Learning Algorithms**

# Naïve Bayes Classifier Algorithm (Supervised Learning - Classification)

The Naïve Bayes classifier is based on Bayes’ theorem and classifies every value as independent of any other value. It allows us to predict a class/category, based on a given set of features, using probability. Despite its simplicity, the classifier does surprisingly well and is often used due to the fact it outperforms more sophisticated classification methods.

# K Means Clustering Algorithm (Unsupervised Learning - Clustering)

The K Means Clustering algorithm is a type of unsupervised learning, which is used to categorise unlabelled data, i.e. data without defined categories or groups. The algorithm works by finding groups within the data, with the number of groups represented by the variable K. It then works iteratively to assign each data point to one of K groups based on the features provided.

# Linear Regression (Supervised Learning/Regression)

Linear regression is the most basic type of regression. Simple linear regression allows us to understand the relationships between two continuous variables.

# Logistic Regression (Supervised learning – Classification)

Logistic regression focuses on estimating the probability of an event occurring based on the previous data provided. It is used to cover a binary dependent variable, that is where only two values, 0 and 1, represent outcomes.

# Artificial Neural Networks (Reinforcement Learning)

An artificial neural network (ANN) comprises ‘units’ arranged in a series of layers, each of which connects to layers on either side. ANNs are inspired by biological systems, such as the brain, and how they process information. ANNs are essentially a large number of interconnected processing elements, working in unison to solve specific problems. ANNs also learn by example and through experience, and they are extremely useful for modelling non-linear relationships in high-dimensional data or where the relationship amongst the input variables is difficult to understand.

# Decision Trees (Supervised Learning – Classification/Regression)

A decision tree is a flow-chart-like tree structure that uses a branching method to illustrate every possible outcome of a decision. Each node within the tree represents a test on a specific variable – and each branch is the outcome of that test.

# Random Forests (Supervised Learning – Classification/Regression)

Random forests or ‘random decision forests’ is an ensemble learning method, combining multiple algorithms to generate better results for classification, regression and other tasks. Each individual classifier is weak, but when combined with others, can produce excellent results. The algorithm starts with a ‘decision tree’ (a tree-like graph or model of decisions) and an input is entered at the top.

# Nearest Neighbours (Supervised Learning)

The K-Nearest-Neighbour algorithm estimates how likely a data point is to be a member of one group or another. It essentially looks at the data points around a single data point to determine what group it is actually in.