Machine Learning Algorithms Overview

*Select one or more choices from the list of common Machine Learning Algorithms, do some investigations and write me a short summary*

Linear Regression

One of the most basic forms of machine learning and it is a type of supervised learning as it uses true labels for training. Supervised learning algorithms have an input variable and output variable.

The model is used to determine a relationship between two variables; a dependant variable changing as the independent variable changed by fitting a linear equation to the data.

To use linear regression the relationship between variables would need to be linear and is often used for predictive analysis, operation efficiency, forecasting/new insights & modelling. Common examples would be forecasting sales of a particular product or comparing blood type against country.

K-means

One of the most popular and simplest forms of unsupervised machine learning as it does have true labels for training.

The model groups similar data points together or into clusters to discover underlying patterns, in simple words, the aim is to segregate groups with similar traits and assign them into clusters.

Examples of when K-means clusters are used are:

Document classification - *cluster documents in multiple categories based on tags, topics, and the content of the document.*

Delivery Store Optimization - *optimize the process of good delivery using truck drones by using a combination of k-means to find the optimal number of launch locations and a genetic algorithm to solve the truck route as a traveling salesman problem.*

Identifying Crime Localities - *data related to crimes available in specific localities in a city, the category of crime, the area of the crime, and the association between the two can give quality insight into crime-prone areas within a city or a locality.*

Customer Segmentation - *helps marketers improve their customer base, work on target areas, and segment customers based on purchase history, interests, or activity monitoring.*

Fraud detection - *utilizing past historical data on fraudulent claims, it is possible to isolate new claims based on its proximity to clusters that indicate fraudulent patterns.*