*Task:*

*Select one or more choices from the list of common Machine Learning Algorithms, do some investigations and write me a short summary. I am looking for the following:*

*• Is it Supervised/Unsupervised/Reinforcement learning?*

*• What does the algorithm do?*

*• In which situations will it be most useful?*

*• (Optional) Can you find any examples of where this algorithm has been used?*

KNN (K-Nearest Neighbours)

The K-Nearest Neighbours is a supervised machine learning algorithm which was developed in the 1950s by Evelyn Fix and Joseph Hodges, (Fix and Hodges, 1951). Machine learning algorithms are defined as a set of instructions that help in problem solving. Supervised machine learning models are particularly used to solve classification or regression problems. As a supervised model, the K-Nearest Neighbours algorithm relies on labelled input data in order to learn a particular function and produce the appropriate output when presented with unlabelled data. Datasets are classed into training and testing data with the training datasets producing classification or prediction outputs, (Dey, 2016).

Machine learning algorithms have a myriad of uses such as data mining, predictive analytics, image processing etc. The K-Nearest Neighbours uses proximity in order to identify similar instances within a dataset and classifies objects based on commonality with those surrounding it. It uses Euclidean distance between the test dataset and training dataset in order to classify the data, (Peterson, 2009). Thus, it’s a model used to solve classification problems especially in cases where the data distribution is known.

The model is one of the most widely used machine learning algorithms due to its simplicity, ease of interpretation, calculation and prediction power. The method is widely applied in economic forecasting, data compression and genetics. It can also be used in political science for predicting voter behaviours or political affiliations. In the banking and finance industry it can be applied in order to predict one’s credit worthiness as well as for calculating credit scores. Organisations such as Amazon and Netflix are also known to use the K-Nearest Neighbours algorithm in predicting consumer behaviours and recommending products and services to users.

Linear Regression

Linear Regression is a supervised machine learning algorithm viewed as one of the most popular choices due to its ease of use, interpretability, scalability and its ability to perform well in online settings, Keboola.com, 2020). It is a predictive model used for exploring the cause-and-effect relationship between an independent variable and a dependent variable with continuous numerical values. It seeks to establish whether there is a linearity in the relationship between the input variable and the output variable. It is used for monitoring any changes in a target variable with respect to the input variable.

Linear Regression is used in solving a plethora of business forecasting problems such as data mining, trend analysis, predicting prices/ costs of a commodity in the future, predicting future revenue figures or for performance comparisons. It is used in cases where the assumption is that the data trend follows a straight line thus it performs well when there is a correlation between variables. Other assumptions also include homoscedasticity, little or no multicollinearity and normal distribution of values.

The model has real life applicability in the medical industry for example where it can be used to monitor the effects of certain drugs on a medical condition. Another common example of linear regression model is the Body Mass Index widely used within biology. The model has also been widely used in marketing, business and finance, agriculture and a whole host of other sectors.