WID Data week 5

**Linear Regression:**

It is a supervised learning method used for forecasting an effect, trend forecasting and strength of predictors

Used in predicting prices at a future point, and sales forecasts amongst others

**Logistic Regression:**

Is a supervised learning method where the outcome is a binary yes/no, pass/fail

Used in medical fields to predict the risk of developing a disease eg:diabetes.

**Decision Tree:**

Is a supervised learning method, and can be used for solving regression and classification problems. It creates a model that can predict the class or value of a target variable by learning inferred from a training data

Used in many datamining software packages including scikit-learn and Microsoft SQL server

**Support Vector Machine (SVM)**

Is a supervised learning method for solving regression and classification problems. Used in facial recognition software, handwriting recognition and bioinformatics for classification of genes.

**Naive Bayes:**

Is a supervised learning method based on Bayes theorem and is better for categorical variable input as opposed to numerical input. It requires less training data than other methods. It is also faster than other methods however its estimations can be wrong in some cases as it assumes all features are independent which rarely happens in real life

It is used for spam detection, and real time predictions

**KNN (K- Nearest Neighbours):**

It is a supervised learning method that assumes that similar things exist in close proximity. It is used in recommender systems on amazon and Netflix etc to direct the customer to other items that they may be interested in. It can be slow to run if there is a lot of data

**K-Means:**

Is an unsupervised learning method used to identify clusters of data objects in a dataset. Can be used to identify insurance fraud detection, optimizing good deliveries and spotting crime localities

**Random Forest**

Is a supervised learning method, it uses labelled data to ‘learn’ to classify unlabelled data, one of the most used algorithms. It is compiled of many decision trees can be used to predict banking fraud and predicting drug sensitivity in the medical field