# Analysis of Breast Cancer Wisconsin (Diagnostic) Data Set

Breast Cancer Wisconsin (Diagnostic) Kaggle dataset is used to determine which characteristics are most useful in predicting whether a cancer is malignant or benign. In the United States most common type of cancer is breast cancer which is nearly one-third of all cancers diagnosed in women and second leading cause of cancer death. Abnormal cell growth in the breast tissue causes Breast Cancer, which is commonly referred to as a Tumour. A tumour does not always indicate cancer; a tumour might be benign (non-cancerous), pre-malignant (cancerous), or malignant (cancerous).The main objective of this analysis is to predict which type of cancer is spreading more rapidly and which feature is more helpful to predict the cancer.

The standard error, mean and worst of these features are used for computation for every image and results in 30 various features. For example, field 3 is Mean Radius, field 13 is Radius SE, and field 23 is Worst Radius.

The following image is data visualization using histogram of all features.
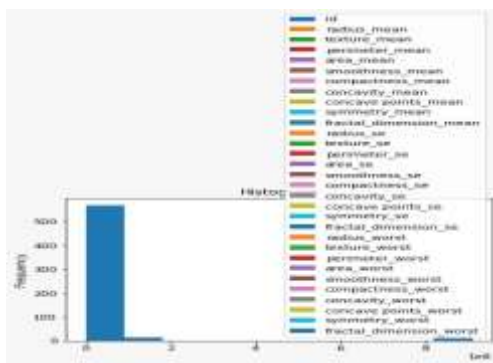

When the data is analysed and observed through bar graph. It shows that the rate of benign tumour is high as compared to malignant tumour.
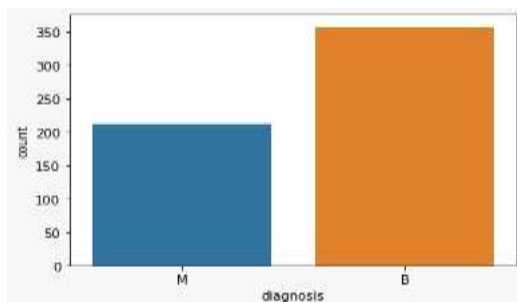
It is noticeable that when the diagnosis is evaluated through scatter plotting. The ratio of malignant is high when measured through radius mean as compared to benign.

Approximately 63% of the observations is indicating that ratio of non-cancerous cell is high where about 37% account cancerous cell in pie chart.
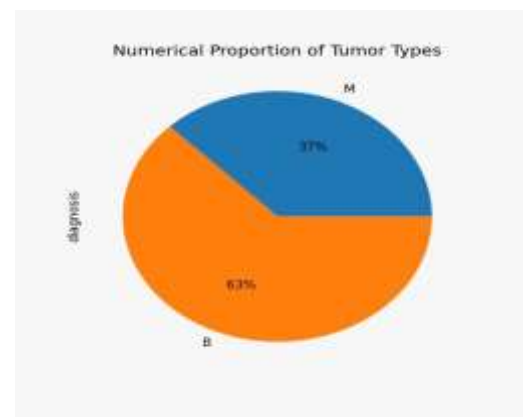
The two important tumour diagnoses that come in result of Wisconsin Dataset are malignant tumour and benign tumour. In case of malignant tumour, the cell starts dividing rapidly and spread all over the body leads to the cancerous cells however the benign tumour is also unusual activity but does not lead to cancerous cells. According to the feature analysis, there are only a few features that have a higher predictive value for the diagnosis.

Dataset from Kaggle.com