**Project: Predicting Breast Cancer**

This project focuses in predicting the probability that a diagnosed breast cancer case is malignant or benign based on Wisconsin dataset from UCI repository.

#### **Dataset and Inputs:**

The dataset I am using in this example is the **Breast Cancer Wisconsin (Diagnostic) Dataset**. The data was downloaded from the [UC Irvine Machine Learning Repository](http://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29).

The input variables which are present in the data are:

* Sample ID (code number)
* Clump thickness
* Uniformity of cell size
* Uniformity of cell shape
* Marginal adhesion
* Single epithelial cell size
* Number of bare nuclei
* Bland chromatin
* Number of normal nuclei
* Mitosis

The target variable which we need to predict is

* Classes

**Techniques applied:**

I have applied various techniques like Logistic regression, Random Forest and Adaboost to predict the target variable.

**Evaluation Metric:**

F1 score is used as evaluation metric to compare the models.

**Conclusion:**

On comparing all the three models, Random forest is giving best results with an F1 score of 0.96