**Project: Breast Cancer Detection**

**Project description:**

Using Machine Learning to Predict Diagnosis of a Breast Cancer

1. Identify the problem

Breast cancer is the most common malignancy among women, accounting for nearly 1 in 3 cancers diagnosed among women in the United States, and it is the second leading cause of cancer death among women. Breast Cancer occurs as a result of abnormal growth of cells in the breast tissue, commonly referred to as a Tumour. A tumor does not mean cancer - tumours can be benign (not cancerous), pre-malignant (pre-cancerous), or malignant (cancerous). Tests such as MRI, mammogram, ultrasound and biopsy are commonly used to diagnose breast cancer performed.

1.1 Expected outcome

Given breast cancer results from breast fine needle aspiration (FNA) test (is a quick and simple procedure to perform, which removes some fluid or cells from a breast lesion or cyst (a lump, sore or swelling) with a fine needle like a blood sample needle). Since this build a model that can classify a breast cancer tumor using two training classification:

1= Malignant (Cancerous) - Present

0= Benign (Not Cancerous) -Absent

1.2 Objective

Since the labels in the data are discrete, the predication falls into two categories, (i.e. Malignant or benign). In machine learning this is a classification problem.

Thus, the goal is to classify whether the breast cancer is benign or malignant and predict the recurrence and non-recurrence of malignant cases after a certain period. To achieve this we have used machine learning classification methods to fit a function that can predict the discrete class of new input.

1.3 Identify data sources.

The Breast Cancer datasets is available machine learning repository maintained by the University of California, Irvine. The dataset contains 569 samples of malignant and benign tumour cells.

The first two columns in the dataset store the unique ID numbers of the samples and the corresponding diagnosis (M=malignant, B=benign), respectively.

The columns 3-32 contain 30 real-value features that have been computed from digitized images of the cell nuclei, which can be used to build a model to predict whether a tumour is benign or malignant.

**Evaluation Scheme:**

**Total marks:** **100**

**Deliverables [Total marks - 95]:**

1. Loading the data into python 🡪 10 marks
2. Performing EDA 🡪 30 marks
3. Training the data by a classification model with >= 80% accuracy 🡪 40 marks
4. Making the predictions for a single test record 🡪 15 marks

**Project Submission [Total marks - 5]:**

1. Once the project has been created, upload all the files on GitHub & commit (save) all the changes, make sure you add a readme file containing detailed description of your thoughts during the project creation. **[3 marks]**
2. Once done, kindly copy the GitHub link of your project & submit the same using your dashboard. **[2 mark]**