**Breast Cancer Prediction**

**By Sana Basharat**

**Project description:** Predicting whether a breast tumor is malignant or benign by training a dataset using Logistic Regression, Random Forest and K Nearest Neighbor Classifiers.

**Project name:** Breast Cancer Prediction using Machine Learning

**Business problem statement:** When health practitioners look at the X-ray scans of breast cancer patients, they have to declare whether the tumor is benign or malignant by considering various factors such as the radius, texture, perimeter, area and smoothness of the tumor as shown in the scans. Because they are humans who are prone to errors, they might make mistakes of false positives or even worse, false negatives.

**AI Business solution description:** We will use AI on a dataset of previous true positives and true negatives, in order to train it to identify whether a tumor is benign or malignant.

**Map a business problem to machine/deep learning solution:** Since a doctor cannot possibly learn previous data that spans across hundreds of samples, we will make a machine look at it and devise patterns so that it can use this data in its predictions.

**Select the right models for a given AI problem:** Training the dataset using three classifiers and comparing their results, implementing LIME to explain the model and deploying through Flask.

**Alternatives considered:** Logistic Regression, Random Forest and K Nearest Neighbor Classifiers

**Decisions made:** Random Forest Classifier turned out to be the best of the three.

**Train and evaluate model:** The 569 sample dataset was split into a ratio of 80:20 for training and testing. The results were explained using LIME.

**Tune model hyper-parameters:** N/A

**Technologies used:** Python, Scikit-learn, LIME, Pickle/Flask

**Results achieved/planned:** Out of the 114 samples used for testing, Random Forest correctly classified 108 of the samples and misclassified the remaining 6.

**Future improvements/releases:** Converting to DL solution