**Classification with Random Forest**

Wisconsin Breast Cancer Machine Learning: [Original Link (Links to an external site.)](https://www.kaggle.com/raviolli77/random-forest-in-python)

<https://www.kaggle.com/code/raviolli77/random-forest-in-python/notebook>

For this project, please implement a **Random Forest Model** on a data set containing descriptive attributes of digitized images of a process known as, *fine needle aspirate* (**FNA**) of breast mass.

We have a total of 29 features that were computed for each cell nucleus with an ID Number and the Diagnosis (Later converted to binary  representations: *Malignant* = **1**, *Benign* = **0**).

This data set originated in early 1990's, when Dr. William H. Wolberg was curious if he could find a way to accurately predict breast cancer diagnosis based on **FNA**'s.

**Tasks:**

1. Implement a Random Forest Model for classification of the FNA images. (5 points)

2. Conduct N-fold cross validation. (1 point)

3. Develop a confusion matrix for each classification result. (1 points)

4. Create ROC curve for training (optional). (1 point bonus)

5. Write a technical report on the model fitting process including probably data cleaning and fixing, bagging, subset choosing, etc, and validation. Submit your Python code and report together. (3 points)

**Data set:**<https://www.kaggle.com/datasets/uciml/breast-cancer-wisconsin-data>

**File:** data-breastCancer.csv