Option B - Breast Cancer Detection using Random Forest Classifier

In the assignment, the breast cancer dataset from sklearn.datasets was used to train a machine learning model for predicting whether a tumor is malignant or benign. We use 5-fold cross-validation to evaluate its performance. To optimize the mode, GRideSEarchCV is used for hyperparameter tuning, focusing on maximizing recall (sensitivity) for medical use cases where false negatives can have serious consequences.

Model Implementation
- **Dataset:** The dataset from *sklearn.datasets* was used, containing 30 features
- **Train-Test Split:** Data was split into **80%** for training and **20%** for testing using *sklearn.model_selection train_test_split* function
- **Cross validation:** A **5-fold** cross validation was performed with *sklearn.model_selection* with KFold
- **Hyperparameter Tuning:** The grid search method was used to optimize the Random Forest Classifier with varying values for *n_estimators, max_depth, criterion*
- **Scoring Metric:** The model was **evaluated using recall (sensitivity)** as the primary metric, since it is **crucial to detect** as **true-positives as possible** in a medical setting.

Performance Metrics / Screenshot:
1. **Accuracy**   - 0.96
2. **Precision**   - 0.96
3. **Recall (Sensitivity)** - 0.97
4. **Specificity**   - 0.93
5. **F1 Score**   - 0.97

```
Prateek Singh@DESKTOP-SANCAM5 MINGW64 /f/Studies/Master's Notes/Spring 2025/SAT 5114 - A.I in Healthcare
$ "C:/Users/Prateek Singh/AppData/Local/Programs/Python/Python310/python.exe" "f:/Studies/Master's Notes/Spring 2025/SAT 5114 - A.I in Healthcare/Assignment 2.py"
Best Score: 0.9785750360750362
Best Hyperparameters: {'criterion': 'entropy', 'max_depth': 5, 'n_estimators': 100}
=================================================
================= Confusion Matrix ==============
=================================================
[[39  3]
 [ 2 70]]
=================================================
================== Model Scores =================
=================================================
           precision    recall  f1-score   support

        0       0.95      0.93      0.94        42
        1       0.96      0.97      0.97        72

 accuracy                           0.96       114
macro avg        0.96      0.95      0.95       114
weighted avg     0.96      0.96      0.96       114

Specificity : 0.93
F-1 Score   : 0.97
```

Conclusion
- **Accuracy of 0.96** and **Recall of 0.97** demonstrate reliable classification and effective identification of malignant tumors.
- **F1-score of 0.97** shows a strong balance between precision and recall, making it effective for breast cancer detection.

Github Link - https://github.com/Prateeek73/SAT5114--Assign2