Fetal Health Classification

By: Reema Alnafisi

Project Components:

1. Design

The model should help the health professionals to investigate the signals resulting from the cardiotogram. This will determine when should they take action according to the fetal and mother health and may help in reducing fetal and maternal mortality.

2. Data

Dataset: The dataset was downloaded from Kaggle and can be found in my github page: https://github.com/reemaaln/bootcamp_project

The used dataset contains 2126 records of features extracted from Cardiotocogram exams, which were then classified by three expert obstetritians into 3 classes:

- Normal
- Suspect
- Pathological

3. Algorithm

Data preprocessing and feature scaling was done on the data. No null or missing values were found.

Then, cross validation score was calculated on different classifiers to predict efficiency of the model. The Random Forest Classifier was chosen based on the results.

The model was built and the hyper parameters were chosen according to GridSearchCV.

Eventually, evaluation metrics were presented. Also, a confusion matrix was illustrated.

- Model Selection cross validation scores:

```
Logistic Regression: 0.897170
Decision Tree: 0.916683
RandomForest: 0.938854
SVC: 0.906594
```

- Model Building

Random Forest:
An ensemble method, meaning that a random forest model is made up of a large number of small decision trees, called estimators, which each produce their own predictions. The random forest model combines the predictions of the estimators to produce a more accurate prediction.

GridSearchCV:

```
#Find the best estimator according to GridSearchCV
best_estimator_RF = CV_RFC.best_params_
print(f"Best estimator for RF: \n{best_estimator_RF}")
```

```
Best estimator for RF:
{'criterion': 'gini', 'min_samples_leaf': 1, 'min_samples_split': 6, 'n_estimators': 500}
```

```
#Find the best score according to GridSearchCV
best_score_RF = CV_RFC.best_score_
print(f"Best score for RF: {round(best_score_RF, 4)}")
```

```
Best score for RF: 0.9388
```

- Evaluation:

```
Mean Square Error= 0.08
Root Mean Square Error= 0.283
R^2 on training set = 0.995
R^2 on testing set = 0.944
```

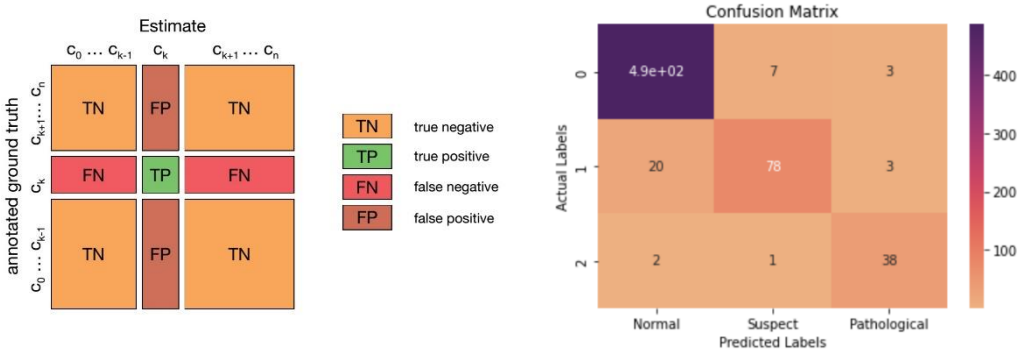
```
Classification Report
              precision    recall  f1-score   support

         1.0       0.96      0.98      0.97       496
         2.0       0.91      0.77      0.83       101
         3.0       0.86      0.93      0.89        41

    accuracy                           0.94       638
   macro avg       0.91      0.89      0.90       638
weighted avg       0.94      0.94      0.94       638
```

We got low errors and high accuracy, precision, f1-score, and recall.

Confusion Matrix:



4. Tools

Libraries needed:

Data Processing: NumPy, Pandas

Modelling: SciKit-Learn

Visualization: Matplotlib, Seaborn

5. Communication

A 5 minutes slides presentation will be presented at the end of the course.