EARIN Lab

Task 4

Variant 3

Group 3

By: Krzysztof Kotowski and Juan Manuel Aristizabal Henao

Task description:

- Create an Iris species classifier
- We are using LogisticRegression and RandomForestClassifier
    - Logistic regression, because it is fit for determining outcomes coming from multiple variables
    - Random forest classifier, because it is very versatile and can handle multiple variables.
- We added manual input for models: LogisticRegression(max_iter) and RandomForestClassifier(rand)
- Training data and test data for model training were split 3 to 1 for training and testing, in that order

Implementation:

- We created a program, which:
    - Reads Iris data
    - Prepares models
    - Tests them using cross validation
- One part takes some sample training and test data for single instances of models, tests them and compares to random guessing
- Second (and third) part checks models for multiple values of their parameters, for Logistic regression it is maximum iterations and for random forest classifier it is number of estimators (trees)

Results:

```
Predictions:
[1 0 2 1 1 0 1 2 1 1 2 0 0 0 0 1 2 1 1 2 0 2 0 2 2 2 2 2 0 0] - True values
[1 0 2 1 1 0 1 2 1 1 2 0 0 0 0 1 2 1 1 2 0 2 0 2 2 2 2 2 0 0] - Logistic Regression
[1 0 2 1 1 0 1 2 1 1 2 0 0 0 0 1 2 1 1 2 0 2 0 2 2 2 2 2 0 0] - Random Forest
[2 0 2 1 0 2 0 2 2 0 0 2 0 2 2 2 1 0 1 0 0 0 1 1 2 1 2 0 1 0] - Random Guessing
Logistic Regression Accuracy: 100.0 %
Random Forest Accuracy: 100.0 %
Random Guessing Accuracy: 33.333333333333 %
```

```
Feature names: ['sepal length (cm)', 'sepal width (cm)', 'petal length (cm)', 'petal width (cm)']
Feature importances: [0.10809762 0.03038681 0.43999397 0.42152159]
```
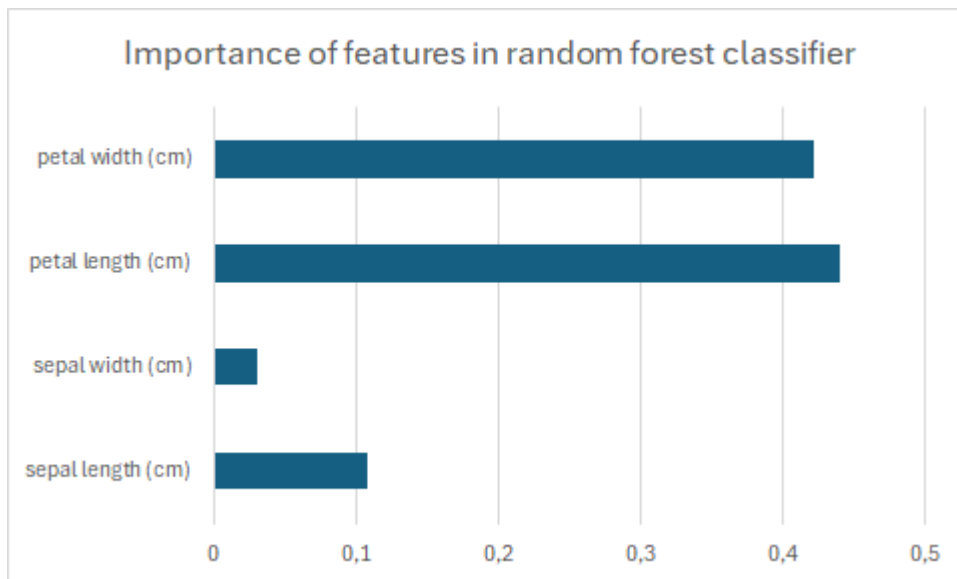

Random forest classifier allowed for checking the importance of each of the parameters on the classification. This is an incredibly useful property of this algorithm for the purposes of this exercise

Random forest prediction accuracy depending on number_of_estimators
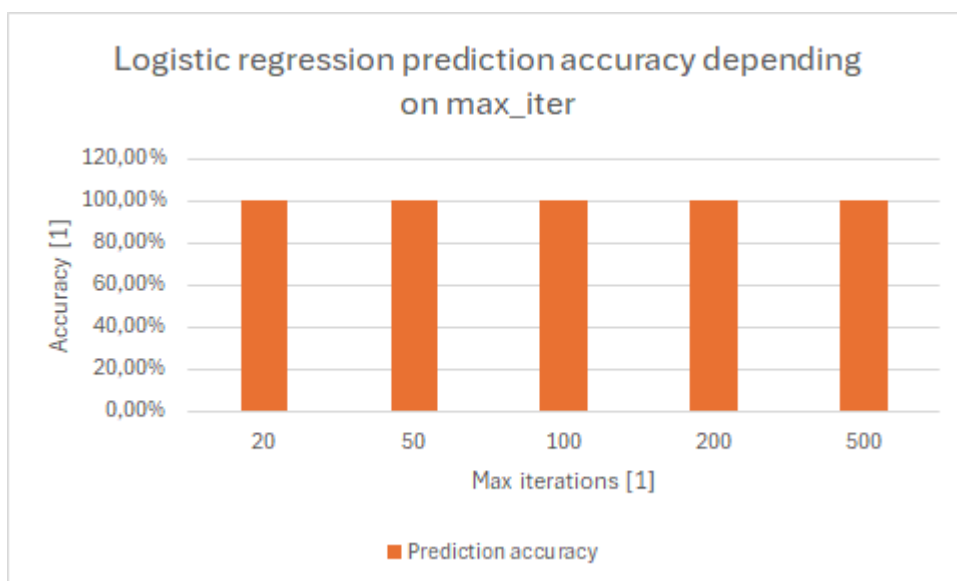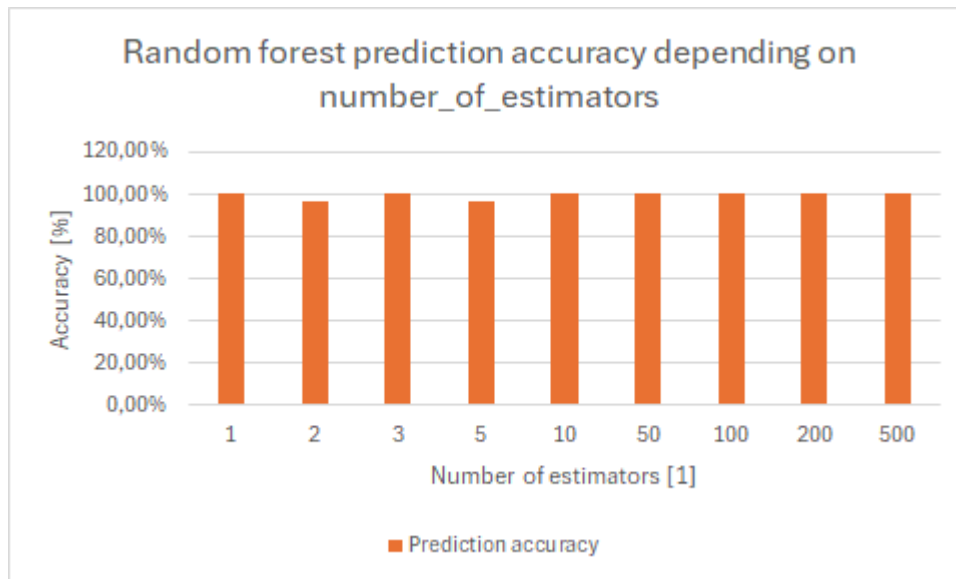
Logistic regression, as well as random forest classifier are a great choice for classifying this data. Answers for test data were almost perfect.