Machine Learning
Report

Task

- Implementation of Naïve Bayes algorithm and the functions to evaluate it with k-fold cross validation.
- Finding the accuracy with code and Weka.
- Comparing both the accuracy obtained with Naïve Bayes and Weka.

Introduction

Naïve Bayes Algorithm:

Naive Bayes uses a similar method to predict the probability of different class based on various attributes. This algorithm is mostly used in text classification and with problems having multiple classes.

The simple form of the calculation for Bayes Theorem is as follows:

$$P(Class/Data) = P(Data/Class) * P(Class)/P(Data)$$

Naïve Bayes is one of the fast and easy ML algorithm to predict a class of datasets. It can be used for binary as well as multi-class Classification. It performs well in multi-class predictions s compared to

the other algorithms. It is the most popular choice for text classification problems.

K-Fold Cross Validation:

K-Fold cross validation is used when the dataset is split into a K number of folds and is used to evaluate the model's ability when given new data. K refers to the number of groups the data sample split into. For example, if you see k value is 10, we can call this as 10-fold cross validation.

Accuracy:

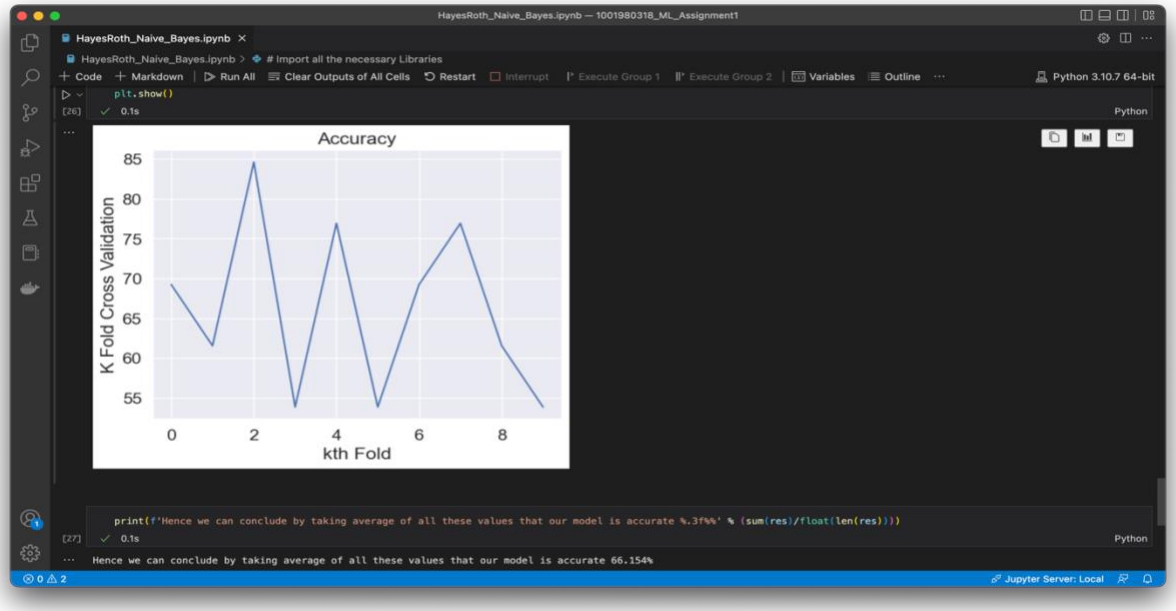
Basically, accuracy is the degree of closeness between measurement and its true value. Here, after we are adding the accuracy to naive bayes algorithm after the 10-fold cross and then check the accuracy of the same dataset using Weka. Then, we compare them.

Accuracy Table:

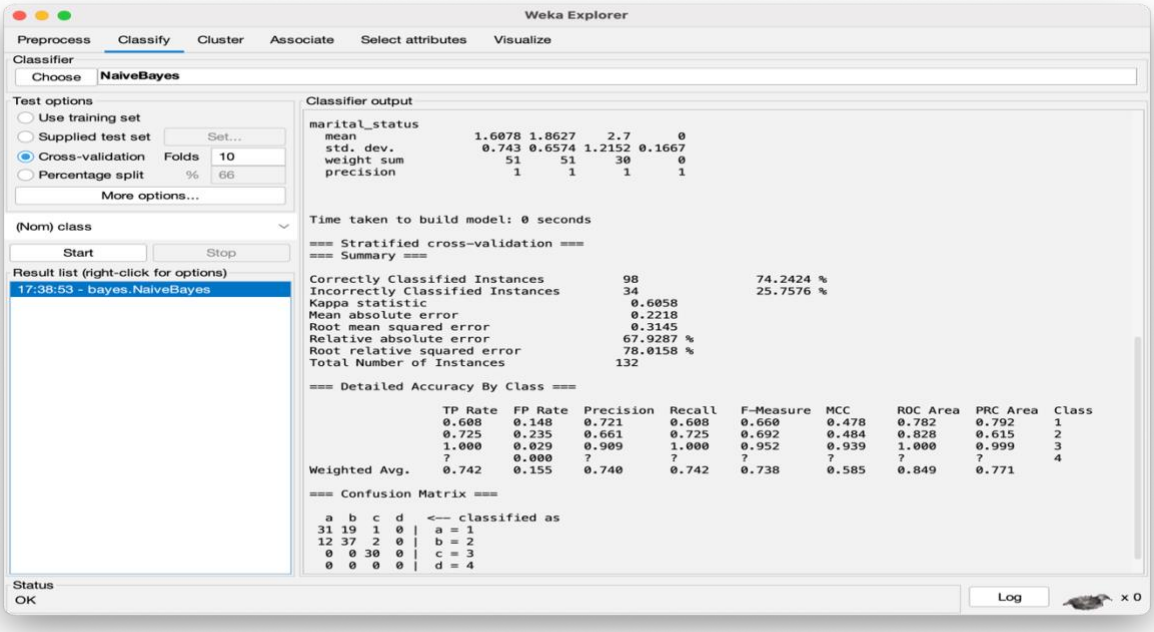
Dataset	Python (%)	Weka (%)
Hayes Roth	66.154	74.242
Breast Cancer	71.852	71.678
Car	73.023	85.532

Python Code and WEKA

Hayes Roth:



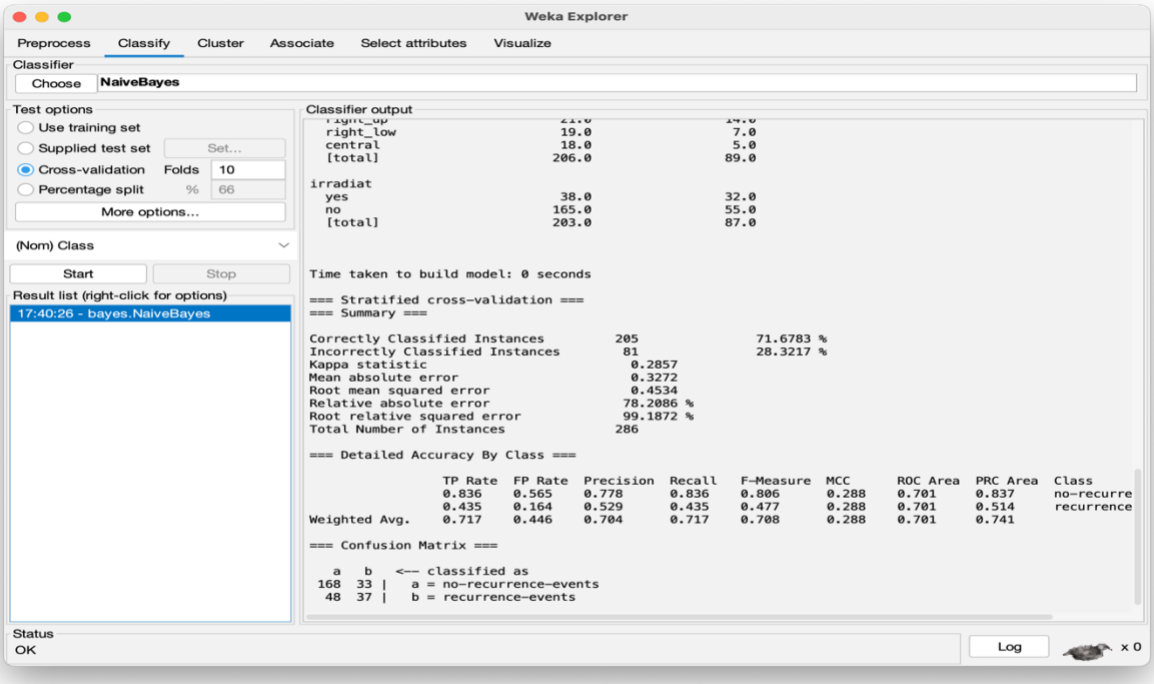
Python



WEKA

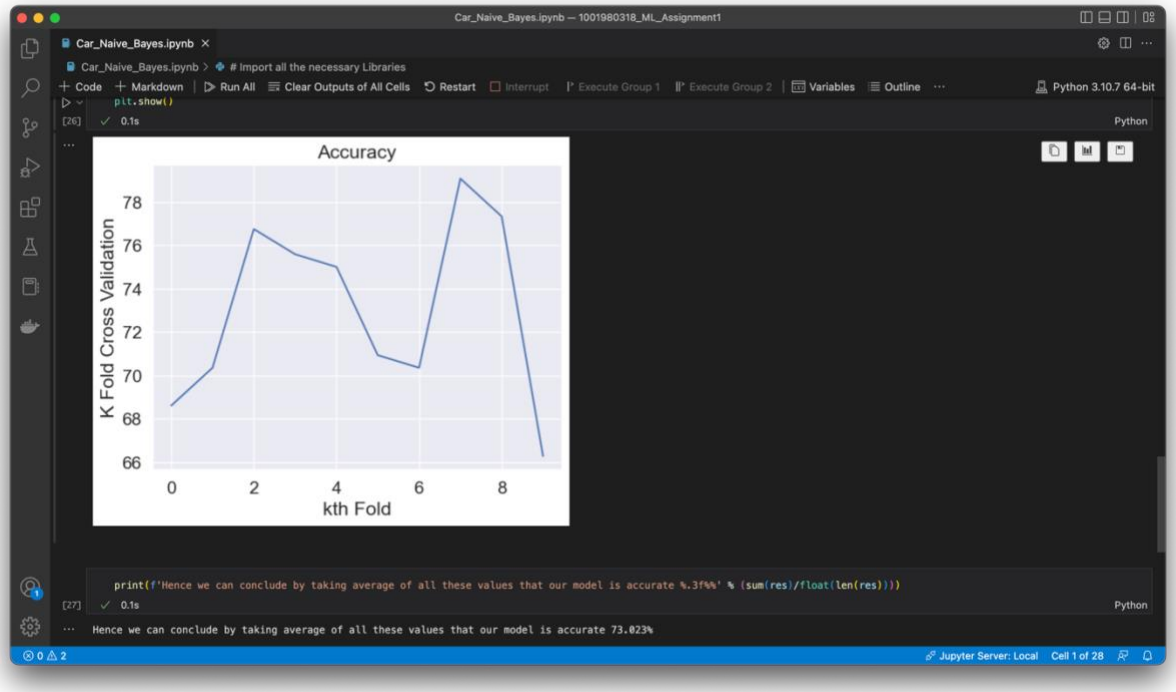
Breast Cancer:

Python

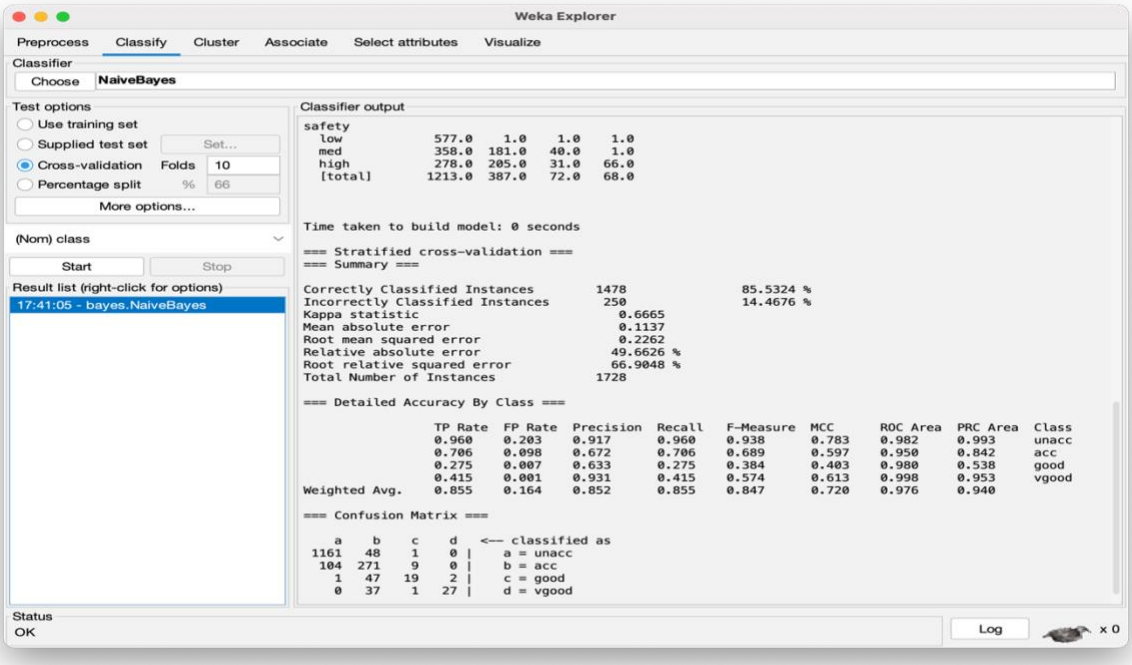


WEKA

Car:



Python



WEKA

References

- https://machinelearningmastery.com/naive-bayes-classifier-scratch-python/
- https://machinelearningmastery.com/load-machine-learning-data-scratch-python/
- https://machinelearningmastery.com/k-fold-cross-validation/
- https://jakevdp.github.io/PythonDataScienceHandbook/05.05-naive-bayes.html
- https://en.wikipedia.org/wiki/Naive_Bayes_classifier