ICP -6 Spark

The focus for implementing the ICP 6 is to execute MLlib is Apache Spark's scalable machine learning library, with APIs in Java, Scala, Python, and R.

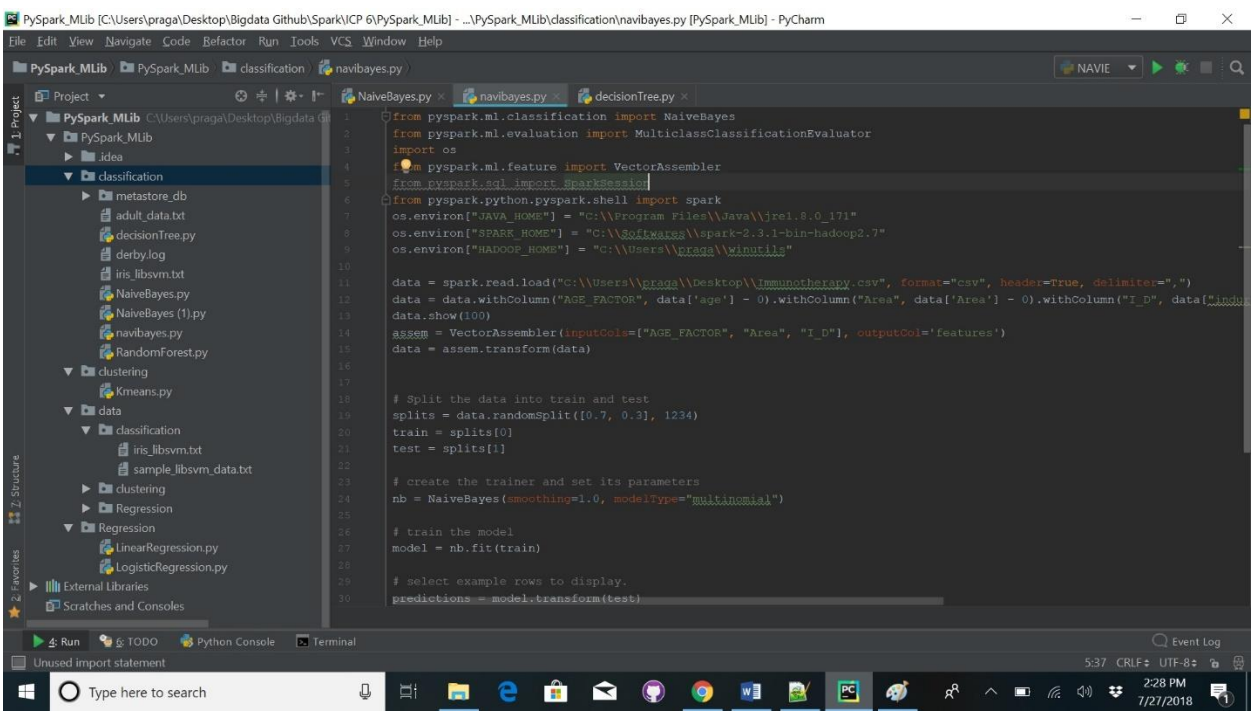
Implemented the **Naive Bayes classifier** on the csv data set. The data set which is used for the execution of the classifier is Immunotherapy.

Columns - sex, age, Time, Number_of_Warts, Type, Area, induration_diameter, (7 - features) Result_of_Treatment(label)

Algorithm: Naive Bayes classifier, In machine learning, naive Bayes classifiers are a family of simple "probabilisticclassifiers" based on applying Bayes' theorem with strong (naive) independence assumptions between the features.

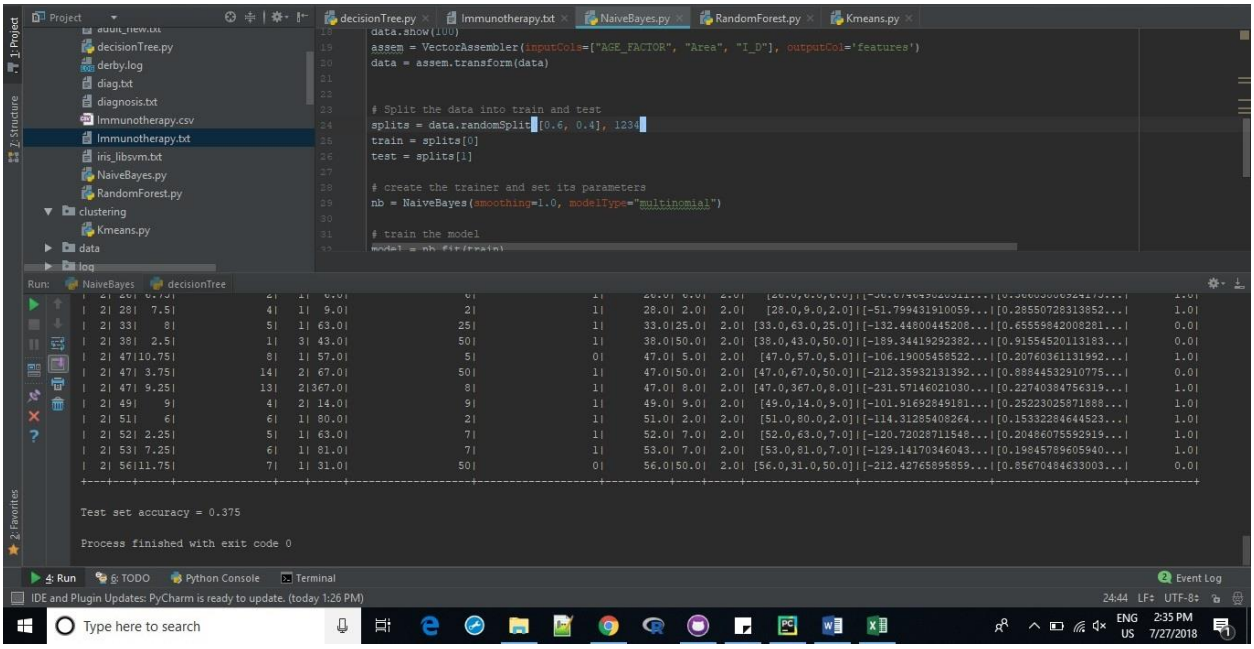
This dataset contains information about wart treatment results of 90 patients using immunotherapy. This dataset has 7 different features like sex, age, time etc., and one label which has two categories (wart treatment curing the disease = 1, wart treatment not curing the disease = 0). In the below code we have only considered three features sex, age and induration_diameter features to predict the Result_of _Treatment. And the cross-validation parameter is set to 60,40 and 90,10 are used to test and train the model.

Code Snippet:



Output Screenshots:

60 and 40 split ratios -Test set accuracy 0.375



70 and 30 split ratios -Test set accuracy -0.4375