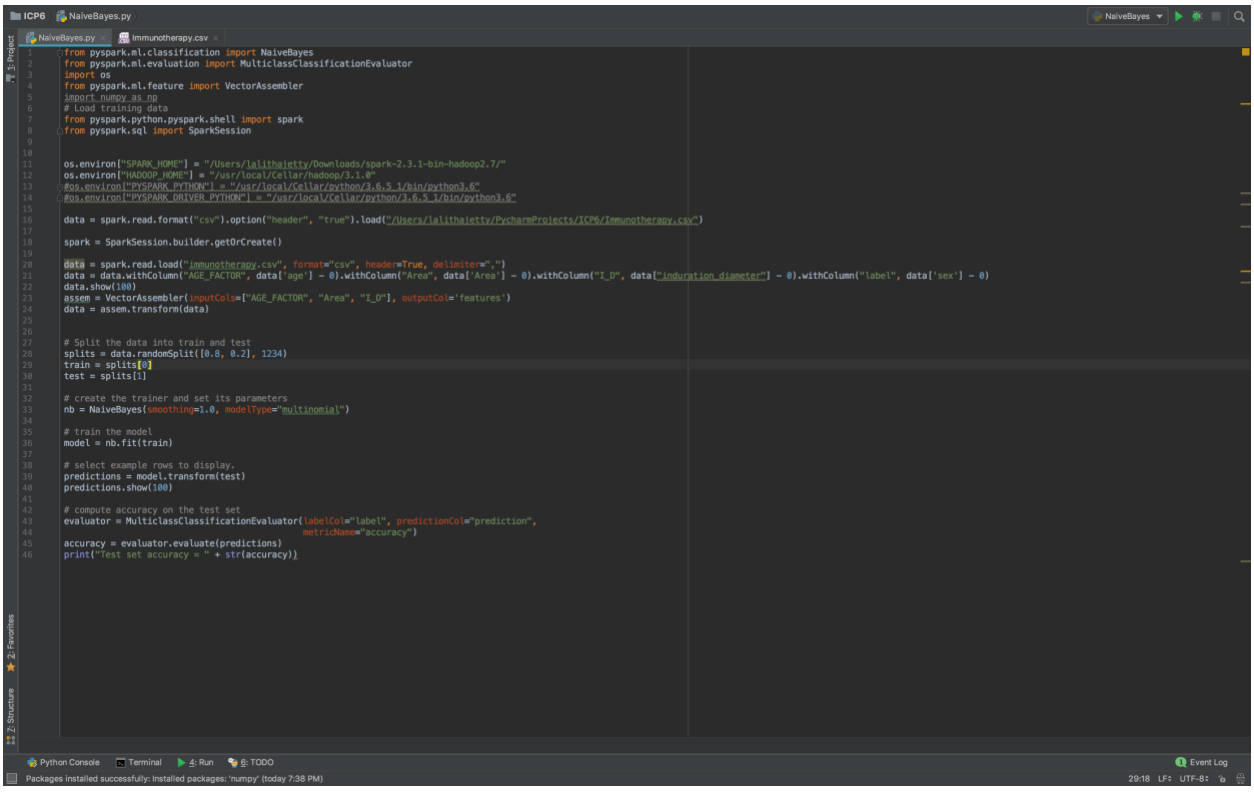
ICP6 – Machine Learning Algorithms

Algorithm – Classification/Naive Bayes
Data Set - Immunotherapy dataset
Columns - sex, age, Time, Number_of_Warts, Type, Area, induration_diameter, (7 - features) Result_of_Treatment(label)

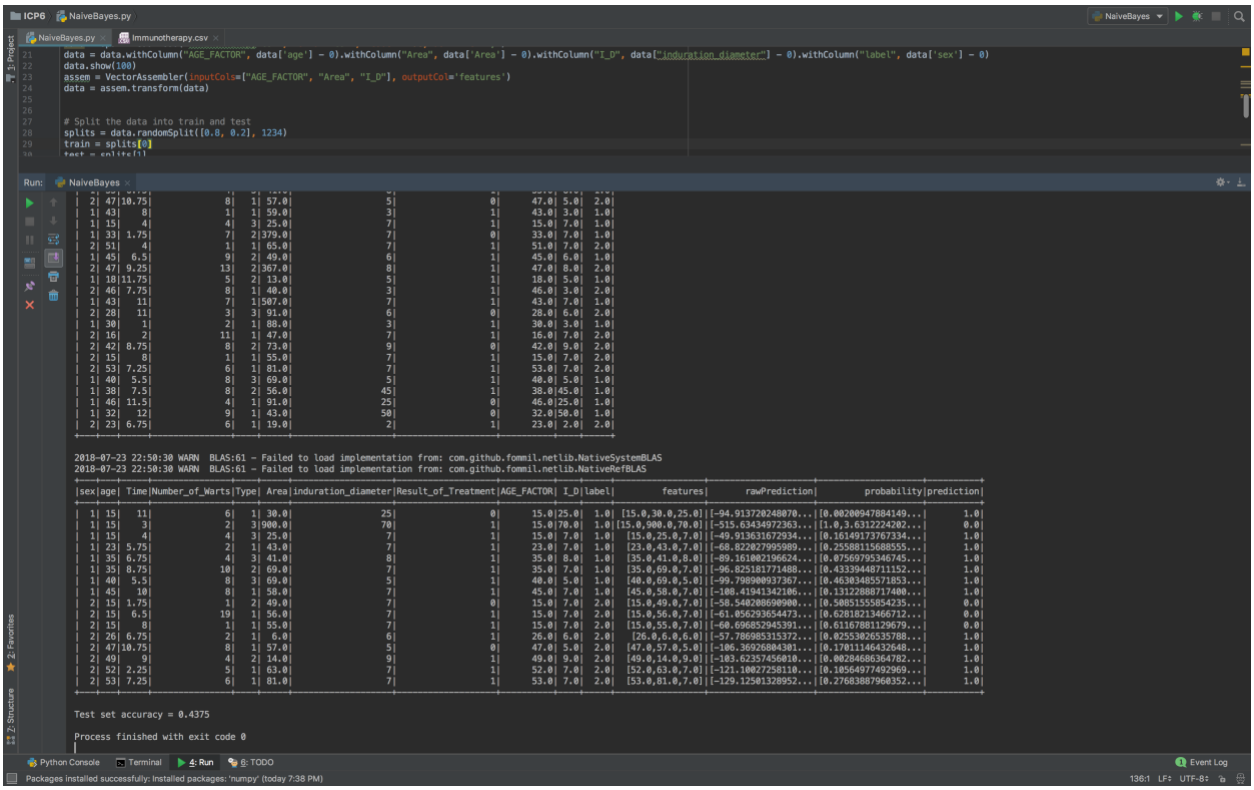
This dataset contains information about wart treatment results of 90 patients using immunotherapy. This dataset has 7 different features like sex, age, time etc., and one label which has two categories (wart treatment curing the disease = 1, wart treatment not curing the disease = 0)

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 80-20 which means 80% of the data = 72 samples are used to train the algorithm and 20% of the data = 18 samples are used to test the model. Random function is used to divide the samples to training and testing. Then the accuracy is calculated which is correct predictions in test set/total predictions in test set.

Code :



Output:



```
data = data.withColumn("AGE_FACTOR", data['age'] - 0).withColumn("Area", data['Area'] - 0).withColumn("I_D", data["induration_diameter"] - 0).withColumn("label", data['sex'] - 0)
data.show(100)
assem = VectorAssembler(inputCols=["AGE_FACTOR", "Area", "I_D"], outputCol='features')
data = assem.transform(data)


# Split the data into train and test
splits = data.randomSplit([0.8, 0.2], 1234)
train = splits[0]
```

```
+---+---+-----+---------------+----+-----+-----------------+------------------+----------+---+-----+
|sex|age| Time|Number_of_Warts|Type| Area|induration_diameter|Result_of_Treatment|AGE_FACTOR|I_D|label|
+---+---+-----+---------------+----+-----+-----------------+------------------+----------+---+-----+
|  1| 47|10.75|              8|   1| 57.0|                5|                 0|      47.0|5.0|  2.0|
|  1| 43|    8|              1|   1| 59.0|                3|                 1|      43.0|3.0|  1.0|
|  1| 15|    4|              4|   3| 25.0|                7|                 1|      15.0|7.0|  1.0|
|  1| 33| 1.75|              7|   2|379.0|                7|                 0|      33.0|7.0|  1.0|
|  2| 51|    4|              1|   1| 65.0|                7|                 1|      51.0|7.0|  2.0|
|  1| 45|  6.5|              9|   2| 49.0|                6|                 1|      45.0|6.0|  1.0|
|  2| 47| 9.25|             13|   2|367.0|                8|                 1|      47.0|8.0|  2.0|
|  1| 18|11.75|              5|   2| 13.0|                5|                 1|      18.0|5.0|  1.0|
|  2| 46| 7.75|              8|   1| 40.0|                3|                 1|      46.0|3.0|  2.0|
|  1| 43|   11|              7|   1|507.0|                7|                 1|      43.0|7.0|  1.0|
|  2| 28|   11|              3|   3| 91.0|                6|                 0|      28.0|6.0|  2.0|
|  1| 30|    1|              2|   1| 80.0|                3|                 1|      30.0|3.0|  1.0|
|  2| 16|    2|             11|   1| 47.0|                7|                 1|      16.0|7.0|  2.0|
|  2| 42| 8.75|              8|   2| 73.0|                9|                 0|      42.0|9.0|  2.0|
|  2| 15|    8|              1|   1| 55.0|                7|                 1|      15.0|7.0|  2.0|
|  2| 53| 7.25|              6|   1| 81.0|                7|                 1|      53.0|7.0|  2.0|
|  1| 40|  5.5|              8|   3| 69.0|                5|                 1|      40.0|5.0|  1.0|
|  1| 38|  7.5|              8|   2| 56.0|               45|                 1|      38.0|45.0|  1.0|
|  1| 46| 11.5|              4|   1| 91.0|               25|                 0|      46.0|25.0|  1.0|
|  1| 32|   12|              9|   1| 43.0|               50|                 0|      32.0|50.0|  1.0|
|  2| 23| 6.75|              6|   1| 19.0|                2|                 1|      23.0| 2.0|  2.0|
+---+---+-----+---------------+----+-----+-----------------+------------------+----------+---+-----+

2018-07-23 22:50:30 WARN  BLAS:61 - Failed to load implementation from: com.github.fommil.netlib.NativeSystemBLAS
2018-07-23 22:50:30 WARN  BLAS:61 - Failed to load implementation from: com.github.fommil.netlib.NativeRefBLAS
+---+---+-----+---------------+----+-----+-----------------+------------------+----------+---+-----+--------------------+--------------------+--------------------+----------+
|sex|age| Time|Number_of_Warts|Type| Area|induration_diameter|Result_of_Treatment|AGE_FACTOR|I_D|label|            features|       rawPrediction|         probability|prediction|
+---+---+-----+---------------+----+-----+-----------------+------------------+----------+---+-----+--------------------+--------------------+--------------------+----------+
|  1| 15|   11|              6|   1| 30.0|               25|                 0|      15.0|25.0|  1.0|[15.0,30.0,25.0]|[-94.913720248070...|[0.00200947884149...|       1.0|
|  1| 15|    3|              2|   3|900.0|               70|                 1|      15.0|70.0|  1.0|[15.0,900.0,70.0]|[-515.63434972363...|[1.0,3.6312224202...|       0.0|
|  1| 15|    4|              4|   3| 25.0|                7|                 1|      15.0| 7.0|  1.0|  [15.0,25.0,7.0]|[-49.913631672934...|[0.16149173767334...|       1.0|
|  1| 23| 5.75|              2|   1| 43.0|                7|                 1|      23.0| 7.0|  1.0|  [23.0,43.0,7.0]|[-68.822027995989...|[0.25588115688555...|       1.0|
|  1| 35| 6.75|              4|   3| 41.0|                8|                 1|      35.0| 8.0|  1.0|  [35.0,41.0,8.0]|[-89.161002196624...|[0.07569795346745...|       1.0|
|  1| 35| 8.75|             10|   2| 69.0|                7|                 1|      35.0| 7.0|  1.0|  [35.0,69.0,7.0]|[-96.825181771488...|[0.43339448711152...|       1.0|
|  1| 40|  5.5|              8|   3| 69.0|                5|                 1|      40.0| 5.0|  1.0|  [40.0,69.0,5.0]|[-99.798900937367...|[0.46303485571853...|       1.0|
|  1| 45|   10|              8|   1| 50.0|                7|                 1|      45.0| 7.0|  1.0|  [45.0,50.0,7.0]|[-108.41941342106...|[0.13122808717400...|       1.0|
|  2| 15| 1.75|              1|   2| 49.0|                7|                 0|      15.0| 7.0|  2.0|  [15.0,49.0,7.0]|[-58.548208690900...|[0.58051555854235...|       0.0|
|  2| 15|  6.5|             19|   1| 56.0|                7|                 1|      15.0| 7.0|  2.0|  [15.0,56.0,7.0]|[-61.056293654473...|[0.62818213466712...|       0.0|
|  2| 15|    8|              1|   1| 55.0|                7|                 1|      15.0| 7.0|  2.0|  [15.0,55.0,7.0]|[-60.696852945391...|[0.61167881129679...|       0.0|
|  2| 26| 6.75|              2|   1|  6.0|                6|                 1|      26.0| 6.0|  2.0|   [26.0,6.0,6.0]|[-57.786985315372...|[0.02553026535788...|       1.0|
|  2| 47|10.75|              8|   1| 57.0|                5|                 0|      47.0| 5.0|  2.0|  [47.0,57.0,5.0]|[-106.36926004381...|[0.17011146432648...|       1.0|
|  2| 49|    9|              4|   2| 14.0|                9|                 1|      49.0| 9.0|  2.0|  [49.0,14.0,9.0]|[-103.62357456010...|[0.00284686364782...|       1.0|
|  2| 52| 2.25|              5|   1| 63.0|                7|                 1|      52.0| 7.0|  2.0|  [52.0,63.0,7.0]|[-121.10027258110...|[0.10564977492969...|       1.0|
|  2| 53| 7.25|              6|   1| 81.0|                7|                 1|      53.0| 7.0|  2.0|  [53.0,81.0,7.0]|[-129.12501328952...|[0.27683887960352...|       1.0|
+---+---+-----+---------------+----+-----+-----------------+------------------+----------+---+-----+--------------------+--------------------+--------------------+----------+

Test set accuracy = 0.4375

Process finished with exit code 0
```

Test set accuracy = 0.4375