Assignment6_Haque

CSC 4760/6760 Big Data Programming

Assignment 6

Due Date: 11:59 pm, April 14, 2021

Problem 1. (100 points)

Please access to the link below and pick one of the classification algorithms except for Naïve Bayes and Random forest classifier.

https://spark.apache.org/docs/latest/ml-classification-regression.html

Please choose one of the three provided datasets and apply your chosen algorithms to it using SparkML

About the datasets:

Iris.csv: the dataset classifies flower species based on their sepal and petal length. There are three classification labels (setosa, versicolor, and virginica)

Breaset_cancer.csv: the dataset classifies breast cancer diagnose based on lab test results. There are two classification labels (M and B)

Wine_quality.csv: the dataset classifies wine quality based on ingredient specification. They are scored from 4-10 (since we should predict between 2-3 labels, you can encode this score to BAD (score < 6) and GOOD (score >= 6).

Report:

Please write a report illustrating your experiments. You must explain:

In the report, you should include the answers to the following questions.

Explanation of the source code

How you clean your data?, How you encode your data?

- There was one column with string values, so this was changed to using indexing. And the ID was dropped, all other columns were transformed into doubles.

```
data = iris.drop("Id")
labelIndexer = StringIndexer(inputCol="Species",
              outputCol="SpeciesIndexed").fit(data)
IndexedDataset = labelIndexer.transform(data).drop("Species")
IndexedDataset.show()

IndexedDataset = IndexedDataset.withColumn("SepalLengthCm",
IndexedDataset["SepalLengthCm"].cast(DoubleType()))
IndexedDataset = IndexedDataset.withColumn("SepalWidthCm",
IndexedDataset["SepalWidthCm"].cast(DoubleType()))
IndexedDataset = IndexedDataset.withColumn("PetalLengthCm",
IndexedDataset["PetalLengthCm"].cast(DoubleType()))
IndexedDataset = IndexedDataset.withColumn("PetalWidthCm",
IndexedDataset["PetalWidthCm"].cast(DoubleType()))
```

```
+-------------+------------+-------------+------------+--------------+
|SepalLengthCm|SepalWidthCm|PetalLengthCm|PetalWidthCm|SpeciesIndexed|
+-------------+------------+-------------+------------+--------------+
|          5.1|         3.5|          1.4|         0.2|           0.0|
|          4.9|         3.0|          1.4|         0.2|           0.0|
|          4.7|         3.2|          1.3|         0.2|           0.0|
|          4.6|         3.1|          1.5|         0.2|           0.0|
|          5.0|         3.6|          1.4|         0.2|           0.0|
|          5.4|         3.9|          1.7|         0.4|           0.0|
|          4.6|         3.4|          1.4|         0.3|           0.0|
|          5.0|         3.4|          1.5|         0.2|           0.0|
|          4.4|         2.9|          1.4|         0.2|           0.0|
|          4.9|         3.1|          1.5|         0.1|           0.0|
|          5.4|         3.7|          1.5|         0.2|           0.0|
|          4.8|         3.4|          1.6|         0.2|           0.0|
|          4.8|         3.0|          1.4|         0.1|           0.0|
|          4.3|         3.0|          1.1|         0.1|           0.0|
|          5.8|         4.0|          1.2|         0.2|           0.0|
|          5.7|         4.4|          1.5|         0.4|           0.0|
|          5.4|         3.9|          1.3|         0.4|           0.0|
|          5.1|         3.5|          1.4|         0.3|           0.0|
|          5.7|         3.8|          1.7|         0.3|           0.0|
|          5.1|         3.8|          1.5|         0.3|           0.0|
+-------------+------------+-------------+------------+--------------+
```

- Next the the proper required features was choses as shown below. And a dataset was assembled using the required features.

```
required_features = ["SepalLengthCm", "SepalWidthCm",
"PetalLengthCm", "PetalWidthCm"]
assembler = VectorAssembler(inputCols =
required_features, outputCol = 'features')
IndexedDataset_features = assembler.transform(IndexedDataset)
```

```
+-------------+------------+-------------+------------+--------------+-----------------+
|SepalLengthCm|SepalWidthCm|PetalLengthCm|PetalWidthCm|SpeciesIndexed|         features|
+-------------+------------+-------------+------------+--------------+-----------------+
|          5.1|         3.5|          1.4|         0.2|           0.0|[5.1,3.5,1.4,0.2]|
|          4.9|         3.0|          1.4|         0.2|           0.0|[4.9,3.0,1.4,0.2]|
|          4.7|         3.2|          1.3|         0.2|           0.0|[4.7,3.2,1.3,0.2]|
|          4.6|         3.1|          1.5|         0.2|           0.0|[4.6,3.1,1.5,0.2]|
|          5.0|         3.6|          1.4|         0.2|           0.0|[5.0,3.6,1.4,0.2]|
|          5.4|         3.9|          1.7|         0.4|           0.0|[5.4,3.9,1.7,0.4]|
|          4.6|         3.4|          1.4|         0.3|           0.0|[4.6,3.4,1.4,0.3]|
|          5.0|         3.4|          1.5|         0.2|           0.0|[5.0,3.4,1.5,0.2]|
|          4.4|         2.9|          1.4|         0.2|           0.0|[4.4,2.9,1.4,0.2]|
|          4.9|         3.1|          1.5|         0.1|           0.0|[4.9,3.1,1.5,0.1]|
|          5.4|         3.7|          1.5|         0.2|           0.0|[5.4,3.7,1.5,0.2]|
|          4.8|         3.4|          1.6|         0.2|           0.0|[4.8,3.4,1.6,0.2]|
|          4.8|         3.0|          1.4|         0.1|           0.0|[4.8,3.0,1.4,0.1]|
|          4.3|         3.0|          1.1|         0.1|           0.0|[4.3,3.0,1.1,0.1]|
|          5.8|         4.0|          1.2|         0.2|           0.0|[5.8,4.0,1.2,0.2]|
|          5.7|         4.4|          1.5|         0.4|           0.0|[5.7,4.4,1.5,0.4]|
|          5.4|         3.9|          1.3|         0.4|           0.0|[5.4,3.9,1.3,0.4]|
|          5.1|         3.5|          1.4|         0.3|           0.0|[5.1,3.5,1.4,0.3]|
|          5.7|         3.8|          1.7|         0.3|           0.0|[5.7,3.8,1.7,0.3]|
|          5.1|         3.8|          1.5|         0.3|           0.0|[5.1,3.8,1.5,0.3]|
+-------------+------------+-------------+------------+--------------+-----------------+
```

What classification label that you choose.

- The classification label choses was the "SpeciesIndexed", this is the index form of "Species"

```
rf = RandomForestClassifier(labelCol='SpeciesIndexed', featuresCol='features', maxDepth=5)
```

What classification algorithm that you choose.

- 2 classification algoritm was used just for testing purposes

    1. RandomForestClassifier

    ```
    rf = RandomForestClassifier(labelCol='SpeciesIndexed', featuresCol='features', maxDepth=5)
    ```

    2. DecisionTreeClassifier

    ```
    dt = DecisionTreeClassifier(labelCol="SpeciesIndexed", featuresCol="features")
    ```

What the accuracy of the algorithm's prediction is.

```
Test Accuracy of Random Forest Classifier = 0.9459459459459459
Test Accuracy of DecisionTreeClassifier = 0.9459459459459459
```

2) Experimental Results: The screenshot of the terminal output of your program. In the output, you need to output the accuracy of your classification. I uploaded the screenshot of my program in iCollege. You may refer to that image.

Code

```
from os import truncate
from pyspark.sql import SparkSession
from pyspark.ml.feature import StringIndexer
from pyspark.ml.feature import VectorAssembler
from pyspark.sql.types import IntegerType
from pyspark.sql.types import DoubleType
from pyspark.ml import Pipeline
from pyspark.ml.classification import FMClassifier
from pyspark.ml.feature import MinMaxScaler, StringIndexer
from pyspark.ml.classification import RandomForestClassifier
from pyspark.ml.classification import NaiveBayes
from pyspark.ml.evaluation import MulticlassClassificationEvaluator
from pyspark.sql.functions import isnull, when, count, col
from pyspark.ml.classification import LogisticRegression
from pyspark.ml.classification import GBTClassifier
from pyspark.ml.classification import DecisionTreeClassifier


spark = SparkSession \
    .builder \
    .appName('Titanic Data') \
    .getOrCreate()

iris = (spark.read
        .format("csv")
        .option('header', 'true')
        .load("/home/nur_haque/HW6/Iris.csv"))
data = iris.drop("Id")
labelIndexer = StringIndexer(inputCol="Species", outputCol="SpeciesIndexed").fit(data)
IndexedDataset = labelIndexer.transform(data).drop("Species")

IndexedDataset.show()

IndexedDataset = IndexedDataset.withColumn("SepalLengthCm", IndexedDataset["SepalLengthCm"].cast(DoubleType()))
IndexedDataset = IndexedDataset.withColumn("SepalWidthCm", IndexedDataset["SepalWidthCm"].cast(DoubleType()))
IndexedDataset = IndexedDataset.withColumn("PetalLengthCm", IndexedDataset["PetalLengthCm"].cast(DoubleType()))
IndexedDataset = IndexedDataset.withColumn("PetalWidthCm", IndexedDataset["PetalWidthCm"].cast(DoubleType()))

required_features = ["SepalLengthCm", "SepalWidthCm", "PetalLengthCm", "PetalWidthCm"]
assembler = VectorAssembler(inputCols = required_features, outputCol = 'features')
IndexedDataset_features = assembler.transform(IndexedDataset)

(trainingData, testData) = IndexedDataset_features.randomSplit([0.7, 0.3])

rf = RandomForestClassifier(labelCol='SpeciesIndexed', featuresCol='features', maxDepth=5)
model = rf.fit(trainingData)
predictions = model.transform(testData)
evaluator = MulticlassClassificationEvaluator(labelCol='SpeciesIndexed', predictionCol='prediction', metricName='accuracy')
accuracy = evaluator.evaluate(predictions)
print('Test Accuracy of Random Forest Classifier =', accuracy)

IndexedDataset_features.show()
dt = DecisionTreeClassifier(labelCol="SpeciesIndexed", featuresCol="features")
model1 = dt.fit(trainingData)
predictions = model1.transform(testData)
evaluator = MulticlassClassificationEvaluator(labelCol='SpeciesIndexed', predictionCol='prediction', metricName='accuracy')
accuracy = evaluator.evaluate(predictions)
print('Test Accuracy of DecisionTreeClassifier =', accuracy)
```

Output

```
Test Accuracy of Random Forest Classifier = 0.9459459459459459
22/07/19 23:41:05 INFO FileSourceStrategy: Pushed Filters:
22/07/19 23:41:05 INFO FileSourceStrategy: Post-Scan Filters:
22/07/19 23:41:05 INFO FileSourceStrategy: Output Data Schema: struct<SepalLengthCm: string, SepalWidthCm: string, PetalLengthCm: string, PetalWidthCm: string, Species: string ... 3 more fields>
22/07/19 23:41:05 INFO CodeGenerator: Code generated in 40.679 ms
22/07/19 23:41:05 INFO MemoryStore: Block broadcast_34 stored as values in memory (estimated size 173.6 KiB, free 433.1 MiB)
22/07/19 23:41:05 INFO MemoryStore: Block broadcast_34_piece0 stored as bytes in memory (estimated size 24.7 KiB, free 433.1 MiB)
22/07/19 23:41:05 INFO BlockManagerInfo: Added broadcast_34_piece0 in memory on 172.30.59.241:41055 (size: 24.7 KiB, free: 434.2 MiB)
22/07/19 23:41:05 INFO SparkContext: Created broadcast 34 from showString at NativeMethodAccessorImpl.java:0
22/07/19 23:41:05 INFO FileSourceScanExec: Planning scan with bin packing, max size: 4194304 bytes, open cost is considered as scanning 4194304 bytes.
22/07/19 23:41:05 INFO SparkContext: Starting job: showString at NativeMethodAccessorImpl.java:0
22/07/19 23:41:05 INFO DAGScheduler: Got job 13 (showString at NativeMethodAccessorImpl.java:0) with 1 output partitions
22/07/19 23:41:05 INFO DAGScheduler: Final stage: ResultStage 21 (showString at NativeMethodAccessorImpl.java:0)
22/07/19 23:41:05 INFO DAGScheduler: Parents of final stage: List()
22/07/19 23:41:05 INFO DAGScheduler: Missing parents: List()
22/07/19 23:41:05 INFO DAGScheduler: Submitting ResultStage 21 (MapPartitionsRDD[77] at showString at NativeMethodAccessorImpl.java:0), which has no missing parents
22/07/19 23:41:05 INFO MemoryStore: Block broadcast_35 stored as values in memory (estimated size 42.4 KiB, free 433.0 MiB)
22/07/19 23:41:05 INFO MemoryStore: Block broadcast_35_piece0 stored as bytes in memory (estimated size 18.7 KiB, free 433.0 MiB)
22/07/19 23:41:05 INFO BlockManagerInfo: Added broadcast_35_piece0 in memory on 172.30.59.241:41055 (size: 18.7 KiB, free: 434.2 MiB)
22/07/19 23:41:05 INFO SparkContext: Created broadcast 35 from broadcast at DAGScheduler.scala:1513
22/07/19 23:41:05 INFO DAGScheduler: Submitting 1 missing tasks from ResultStage 21 (MapPartitionsRDD[77] at showString at NativeMethodAccessorImpl.java:0) (first 15 tasks are for partitions Vec
22/07/19 23:41:05 INFO TaskSchedulerImpl: Adding task set 21.0 with 1 tasks resource profile 0
22/07/19 23:41:05 INFO BlockManagerInfo: Removed broadcast_8_piece0 on 172.30.59.241:41055 in memory (size: 24.7 KiB, free: 434.2 MiB)
22/07/19 23:41:05 INFO TaskSetManager: Starting task 0.0 in stage 21.0 (TID 20) (172.30.59.241, executor driver, partition 0, PROCESS_LOCAL, 4906 bytes) taskResourceAssignments Map()
22/07/19 23:41:05 INFO Executor: Running task 0.0 in stage 21.0 (TID 20)
22/07/19 23:41:05 INFO BlockManagerInfo: Removed broadcast_33_piece0 on 172.30.59.241:41055 in memory (size: 2.8 KiB, free: 434.2 MiB)
22/07/19 23:41:05 INFO BlockManagerInfo: Removed broadcast_32_piece0 on 172.30.59.241:41055 in memory (size: 53.2 KiB, free: 434.2 MiB)
22/07/19 23:41:05 INFO BlockManagerInfo: Removed broadcast_31_piece0 on 172.30.59.241:41055 in memory (size: 24.7 KiB, free: 434.2 MiB)
22/07/19 23:41:05 INFO FileScanRDD: Reading File path: file:///home/nur_haque/HW6/Iris.csv, range: 0-5107, partition values: [empty row]
22/07/19 23:41:05 INFO BlockManager: Removing RDD 42
22/07/19 23:41:05 INFO BlockManagerInfo: Removed broadcast_28_piece0 on 172.30.59.241:41055 in memory (size: 35.2 KiB, free: 434.3 MiB)
22/07/19 23:41:05 INFO Executor: Finished task 0.0 in stage 21.0 (TID 20). 2055 bytes result sent to driver
22/07/19 23:41:05 INFO TaskSetManager: Finished task 0.0 in stage 21.0 (TID 20) in 67 ms on 172.30.59.241 (executor driver) (1/1)
22/07/19 23:41:05 INFO TaskSchedulerImpl: Removed TaskSet 21.0, whose tasks have all completed, from pool
22/07/19 23:41:05 INFO BlockManagerInfo: Removed broadcast_9_piece0 on 172.30.59.241:41055 in memory (size: 24.7 KiB, free: 434.3 MiB)
22/07/19 23:41:05 INFO DAGScheduler: ResultStage 21 (showString at NativeMethodAccessorImpl.java:0) finished in 0.086 s
22/07/19 23:41:05 INFO DAGScheduler: Job 13 is finished. Cancelling potential speculative or zombie tasks for this job
22/07/19 23:41:05 INFO TaskSchedulerImpl: Killing all running tasks in stage 21: Stage finished
22/07/19 23:41:05 INFO DAGScheduler: Job 13 finished: showString at NativeMethodAccessorImpl.java:0, took 0.093054 s
22/07/19 23:41:05 INFO BlockManagerInfo: Removed broadcast_29_piece0 on 172.30.59.241:41055 in memory (size: 4.3 KiB, free: 434.3 MiB)
22/07/19 23:41:05 INFO CodeGenerator: Code generated in 18.3615 ms
+-------------+------------+-------------+------------+--------------+----------------+
|SepalLengthCm|SepalWidthCm|PetalLengthCm|PetalWidthCm|SpeciesIndexed|        features|
+-------------+------------+-------------+------------+--------------+----------------+
|          5.1|         3.5|          1.4|         0.2|           0.0|[5.1,3.5,1.4,0.2]|
|          4.9|         3.0|          1.4|         0.2|           0.0|[4.9,3.0,1.4,0.2]|
|          4.7|         3.2|          1.3|         0.2|           0.0|[4.7,3.2,1.3,0.2]|
|          4.6|         3.1|          1.5|         0.2|           0.0|[4.6,3.1,1.5,0.2]|
|          5.0|         3.6|          1.4|         0.2|           0.0|[5.0,3.6,1.4,0.2]|
|          5.4|         3.9|          1.7|         0.4|           0.0|[5.4,3.9,1.7,0.4]|
|          4.6|         3.4|          1.4|         0.3|           0.0|[4.6,3.4,1.4,0.3]|
|          5.0|         3.4|          1.5|         0.2|           0.0|[5.0,3.4,1.5,0.2]|
|          4.4|         2.9|          1.4|         0.2|           0.0|[4.4,2.9,1.4,0.2]|
|          4.9|         3.1|          1.5|         0.1|           0.0|[4.9,3.1,1.5,0.1]|
|          5.4|         3.7|          1.5|         0.2|           0.0|[5.4,3.7,1.5,0.2]|
|          4.8|         3.4|          1.6|         0.2|           0.0|[4.8,3.4,1.6,0.2]|
|          4.8|         3.0|          1.4|         0.1|           0.0|[4.8,3.0,1.4,0.1]|
|          4.3|         3.0|          1.1|         0.1|           0.0|[4.3,3.0,1.1,0.1]|
|          5.8|         4.0|          1.2|         0.2|           0.0|[5.8,4.0,1.2,0.2]|
|          5.7|         4.4|          1.5|         0.4|           0.0|[5.7,4.4,1.5,0.4]|
|          5.4|         3.9|          1.3|         0.4|           0.0|[5.4,3.9,1.3,0.4]|
|          5.1|         3.5|          1.4|         0.3|           0.0|[5.1,3.5,1.4,0.3]|
|          5.7|         3.8|          1.7|         0.3|           0.0|[5.7,3.8,1.7,0.3]|
```

```
|          5.1|          3.8|          1.5|         0.3|         0.0|[5.1,3.8,1.5,0.3]|
+-------------+-------------+-------------+------------+-------------+----------------+
only showing top 20 rows

22/07/19 23:41:05 INFO Instrumentation: [dbfe5743] Stage class: DecisionTreeClassifier
22/07/19 23:41:05 INFO Instrumentation: [dbfe5743] Stage uid: DecisionTreeClassifier_d931d9a9210e
22/07/19 23:41:05 INFO FileSourceStrategy: Pushed Filters:
22/07/19 23:41:05 INFO FileSourceStrategy: Post-Scan Filters:
22/07/19 23:41:05 INFO FileSourceStrategy: Output Data Schema: struct<SepalLengthCm:
string, SepalWidthCm: string, PetalLengthCm: string, PetalWidthCm: string, Species: string ... 3 more fields>
22/07/19 23:41:05 INFO MemoryStore: Block broadcast_36 stored as values in memory (estimated size 173.6 KiB, free 433.8 MiB)
22/07/19 23:41:05 INFO MemoryStore: Block broadcast_36_piece0 stored as bytes in memory (estimated size 24.7 KiB, free 433.8 MiB)
22/07/19 23:41:05 INFO BlockManagerInfo: Added broadcast_36_piece0 in memory on 172.30.59.241:41055 (size: 24.7 KiB, free: 434.3 MiB)
22/07/19 23:41:05 INFO SparkContext: Created broadcast 36 from rdd at Instrumentation.scala:62
22/07/19 23:41:05 INFO FileSourceScanExec: Planning scan with bin packing, max size: 4194304 bytes, open cost is considered as scanning 4194304 bytes.
22/07/19 23:41:05 INFO Instrumentation: [dbfe5743] training: numPartitions=1 storageLevel=StorageLevel(1 replicas)
22/07/19 23:41:05 INFO FileSourceStrategy: Pushed Filters:
22/07/19 23:41:05 INFO FileSourceStrategy: Post-Scan Filters:
22/07/19 23:41:05 INFO FileSourceStrategy: Output Data Schema: struct<SepalLengthCm: string,
 SepalWidthCm: string, PetalLengthCm: string, PetalWidthCm: string, Species: string ... 3 more fields>
22/07/19 23:41:05 INFO MemoryStore: Block broadcast_37 stored as values in memory (estimated size 173.6 KiB, free 433.6 MiB)
22/07/19 23:41:05 INFO MemoryStore: Block broadcast_37_piece0 stored as bytes in memory (estimated size 24.7 KiB, free 433.6 MiB)
22/07/19 23:41:05 INFO BlockManagerInfo: Added broadcast_37_piece0 in memory on 172.30.59.241:41055 (size: 24.7 KiB, free: 434.3 MiB)
22/07/19 23:41:05 INFO SparkContext: Created broadcast 37 from rdd at Predictor.scala:81
22/07/19 23:41:05 INFO FileSourceScanExec: Planning scan with bin packing, max size: 4194304 bytes, open cost is considered as scanning 4194304 bytes.
22/07/19 23:41:05 INFO Instrumentation: [dbfe5743] {"numClasses":3}
22/07/19 23:41:05 INFO Instrumentation: [dbfe5743] {"labelCol":"SpeciesIndexed","featuresCol":"features"}
22/07/19 23:41:05 INFO SparkContext: Starting job: take at DecisionTreeMetadata.scala:119
22/07/19 23:41:05 INFO DAGScheduler: Got job 14 (take at DecisionTreeMetadata.scala:119) with 1 output partitions
22/07/19 23:41:05 INFO DAGScheduler: Final stage: ResultStage 22 (take at DecisionTreeMetadata.scala:119)
22/07/19 23:41:05 INFO DAGScheduler: Parents of final stage: List()
22/07/19 23:41:05 INFO DAGScheduler: Missing parents: List()
22/07/19 23:41:05 INFO DAGScheduler: Submitting ResultStage 22 (MapPartitionsRDD[93] at map at DecisionTreeMetadata.scala:119), which has no missing parents
22/07/19 23:41:05 INFO MemoryStore: Block broadcast_38 stored as values in memory (estimated size 65.9 KiB, free 433.5 MiB)
22/07/19 23:41:05 INFO MemoryStore: Block broadcast_38_piece0 stored as bytes in memory (estimated size 26.2 KiB, free 433.5 MiB)
22/07/19 23:41:05 INFO BlockManagerInfo: Added broadcast_38_piece0 in memory on 172.30.59.241:41055 (size: 26.2 KiB, free: 434.3 MiB)
22/07/19 23:41:05 INFO SparkContext: Created broadcast 38 from broadcast at DAGScheduler.scala:1513
22/07/19 23:41:05 INFO DAGScheduler: Submitting 1 missing tasks from ResultStage 22 (MapPartitionsRDD[93]
 at map at DecisionTreeMetadata.scala:119) (first 15 tasks are for partitions Vector(0))
22/07/19 23:41:05 INFO TaskSchedulerImpl: Adding task set 22.0 with 1 tasks resource profile 0
22/07/19 23:41:05 INFO TaskSetManager: Starting task 0.0 in stage 22.0 (TID 21) (172.30.59.241,
executor driver, partition 0, PROCESS_LOCAL, 4906 bytes) taskResourceAssignments Map()
22/07/19 23:41:05 INFO Executor: Running task 0.0 in stage 22.0 (TID 21)
22/07/19 23:41:05 INFO FileScanRDD: Reading File path: file:///home/nur_haque/HW6/Iris.csv, range: 0-5107, partition values: [empty row]
22/07/19 23:41:05 INFO Executor: Finished task 0.0 in stage 22.0 (TID 21). 1798 bytes result sent to driver
22/07/19 23:41:05 INFO TaskSetManager: Finished task 0.0 in stage 22.0 (TID 21) in 44 ms on 172.30.59.241 (executor driver) (1/1)
22/07/19 23:41:05 INFO TaskSchedulerImpl: Removed TaskSet 22.0, whose tasks have all completed, from pool
22/07/19 23:41:05 INFO DAGScheduler: ResultStage 22 (take at DecisionTreeMetadata.scala:119) finished in 0.054 s
22/07/19 23:41:05 INFO DAGScheduler: Job 14 is finished. Cancelling potential speculative or zombie tasks for this job
22/07/19 23:41:05 INFO TaskSchedulerImpl: Killing all running tasks in stage 22: Stage finished
22/07/19 23:41:05 INFO DAGScheduler: Job 14 finished: take at DecisionTreeMetadata.scala:119, took 0.060502 s
22/07/19 23:41:05 INFO SparkContext: Starting job: aggregate at DecisionTreeMetadata.scala:125
22/07/19 23:41:05 INFO DAGScheduler: Got job 15 (aggregate at DecisionTreeMetadata.scala:125) with 1 output partitions
22/07/19 23:41:05 INFO DAGScheduler: Final stage: ResultStage 23 (aggregate at DecisionTreeMetadata.scala:125)
22/07/19 23:41:05 INFO DAGScheduler: Parents of final stage: List()
22/07/19 23:41:05 INFO DAGScheduler: Missing parents: List()
22/07/19 23:41:05 INFO DAGScheduler: Submitting ResultStage 23 (MapPartitionsRDD[92] at retag at RandomForest.scala:274), which has no missing parents
22/07/19 23:41:05 INFO MemoryStore: Block broadcast_39 stored as values in memory (estimated size 65.9 KiB, free 433.4 MiB)
22/07/19 23:41:05 INFO MemoryStore: Block broadcast_39_piece0 stored as bytes in memory (estimated size 26.2 KiB, free 433.4 MiB)
22/07/19 23:41:05 INFO BlockManagerInfo: Added broadcast_39_piece0 in memory on 172.30.59.241:41055 (size: 26.2 KiB, free: 434.2 MiB)
22/07/19 23:41:05 INFO SparkContext: Created broadcast 39 from broadcast at DAGScheduler.scala:1513
22/07/19 23:41:05 INFO DAGScheduler: Submitting 1 missing tasks from ResultStage 23
(MapPartitionsRDD[92] at retag at RandomForest.scala:274) (first 15 tasks are for partitions Vector(0))
22/07/19 23:41:05 INFO TaskSchedulerImpl: Adding task set 23.0 with 1 tasks resource profile 0
22/07/19 23:41:05 INFO TaskSetManager: Starting task 0.0 in stage 23.0 (TID 22) (172.30.59.241,
executor driver, partition 0, PROCESS_LOCAL, 4906 bytes) taskResourceAssignments Map()
22/07/19 23:41:05 INFO Executor: Running task 0.0 in stage 23.0 (TID 22)
22/07/19 23:41:05 INFO FileScanRDD: Reading File path: file:///home/nur_haque/HW6/Iris.csv, range: 0-5107, partition values: [empty row]
22/07/19 23:41:05 INFO Executor: Finished task 0.0 in stage 23.0 (TID 22). 1913 bytes result sent to driver
22/07/19 23:41:05 INFO TaskSetManager: Finished task 0.0 in stage 23.0 (TID 22) in 47 ms on 172.30.59.241 (executor driver) (1/1)
22/07/19 23:41:05 INFO TaskSchedulerImpl: Removed TaskSet 23.0, whose tasks have all completed, from pool
22/07/19 23:41:05 INFO DAGScheduler: ResultStage 23 (aggregate at DecisionTreeMetadata.scala:125) finished in 0.056 s
22/07/19 23:41:05 INFO DAGScheduler: Job 15 is finished. Cancelling potential speculative or zombie tasks for this job
22/07/19 23:41:05 INFO TaskSchedulerImpl: Killing all running tasks in stage 23: Stage finished
22/07/19 23:41:05 INFO DAGScheduler: Job 15 finished: aggregate at DecisionTreeMetadata.scala:125, took 0.062656 s
22/07/19 23:41:06 INFO SparkContext: Starting job: collectAsMap at RandomForest.scala:1054
22/07/19 23:41:06 INFO DAGScheduler: Registering RDD 95 (flatMap at RandomForest.scala:1039) as input to shuffle 8
22/07/19 23:41:06 INFO DAGScheduler: Got job 16 (collectAsMap at RandomForest.scala:1054) with 1 output partitions
22/07/19 23:41:06 INFO DAGScheduler: Final stage: ResultStage 25 (collectAsMap at RandomForest.scala:1054)
22/07/19 23:41:06 INFO DAGScheduler: Parents of final stage: List(ShuffleMapStage 24)
22/07/19 23:41:06 INFO DAGScheduler: Missing parents: List(ShuffleMapStage 24)
22/07/19 23:41:06 INFO DAGScheduler: Submitting ShuffleMapStage 24 (MapPartitionsRDD[95] at flatMap at RandomForest.scala:1039), which has no missing parents
22/07/19 23:41:06 INFO MemoryStore: Block broadcast_40 stored as values in memory (estimated size 70.1 KiB, free 433.3 MiB)
22/07/19 23:41:06 INFO MemoryStore: Block broadcast_40_piece0 stored as bytes in memory (estimated size 27.9 KiB, free 433.3 MiB)
22/07/19 23:41:06 INFO BlockManagerInfo: Added broadcast_40_piece0 in memory on 172.30.59.241:41055 (size: 27.9 KiB, free: 434.2 MiB)
22/07/19 23:41:06 INFO SparkContext: Created broadcast 40 from broadcast at DAGScheduler.scala:1513
22/07/19 23:41:06 INFO DAGScheduler: Submitting 1 missing tasks from ShuffleMapStage 24
(MapPartitionsRDD[95] at flatMap at RandomForest.scala:1039) (first 15 tasks are for partitions Vector(0))
22/07/19 23:41:06 INFO TaskSchedulerImpl: Adding task set 24.0 with 1 tasks resource profile 0
22/07/19 23:41:06 INFO TaskSetManager: Starting task 0.0 in stage 24.0
 (TID 23) (172.30.59.241, executor driver, partition 0, PROCESS_LOCAL, 4895 bytes) taskResourceAssignments Map()
22/07/19 23:41:06 INFO Executor: Running task 0.0 in stage 24.0 (TID 23)
22/07/19 23:41:06 INFO FileScanRDD: Reading File path: file:///home/nur_haque/HW6/Iris.csv, range: 0-5107, partition values: [empty row]
22/07/19 23:41:06 INFO Executor: Finished task 0.0 in stage 24.0 (TID 23). 2093 bytes result sent to driver
22/07/19 23:41:06 INFO TaskSetManager: Finished task 0.0 in stage 24.0 (TID 23) in 48 ms on 172.30.59.241 (executor driver) (1/1)
22/07/19 23:41:06 INFO TaskSchedulerImpl: Removed TaskSet 24.0, whose tasks have all completed, from pool
22/07/19 23:41:06 INFO DAGScheduler: ShuffleMapStage 24 (flatMap at RandomForest.scala:1039) finished in 0.060 s
22/07/19 23:41:06 INFO DAGScheduler: looking for newly runnable stages
22/07/19 23:41:06 INFO DAGScheduler: running: Set()
22/07/19 23:41:06 INFO DAGScheduler: waiting: Set(ResultStage 25)
22/07/19 23:41:06 INFO DAGScheduler: failed: Set()
22/07/19 23:41:06 INFO DAGScheduler: Submitting ResultStage 25 (MapPartitionsRDD[97] at map at RandomForest.scala:1054), which has no missing parents
22/07/19 23:41:06 INFO MemoryStore: Block broadcast_41 stored as values in memory (estimated size 9.7 KiB, free 433.3 MiB)
22/07/19 23:41:06 INFO MemoryStore: Block broadcast_41_piece0 stored as bytes in memory (estimated size 4.5 KiB, free 433.3 MiB)
22/07/19 23:41:06 INFO BlockManagerInfo: Added broadcast_41_piece0 in memory on 172.30.59.241:41055 (size: 4.5 KiB, free: 434.2 MiB)
22/07/19 23:41:06 INFO SparkContext: Created broadcast 41 from broadcast at DAGScheduler.scala:1513
22/07/19 23:41:06 INFO BlockManagerInfo: Removed broadcast_35_piece0 on 172.30.59.241:41055 in memory (size: 18.7 KiB, free: 434.2 MiB)
22/07/19 23:41:06 INFO DAGScheduler: Submitting 1 missing tasks from ResultStage 25
(MapPartitionsRDD[97] at map at RandomForest.scala:1054) (first 15 tasks are for partitions Vector(0))
22/07/19 23:41:06 INFO TaskSchedulerImpl: Adding task set 25.0 with 1 tasks resource profile 0
22/07/19 23:41:06 INFO TaskSetManager: Starting task 0.0 in stage 25.0
 (TID 24) (172.30.59.241, executor driver, partition 0, NODE_LOCAL, 4271 bytes) taskResourceAssignments Map()
22/07/19 23:41:06 INFO Executor: Running task 0.0 in stage 25.0 (TID 24)
22/07/19 23:41:06 INFO BlockManagerInfo: Removed broadcast_34_piece0 on 172.30.59.241:41055 in memory (size: 24.7 KiB, free: 434.2 MiB)
22/07/19 23:41:06 INFO BlockManagerInfo: Removed broadcast_39_piece0 on 172.30.59.241:41055 in memory (size: 26.2 KiB, free: 434.3 MiB)
22/07/19 23:41:06 INFO BlockManagerInfo: Removed broadcast_38_piece0 on 172.30.59.241:41055 in memory (size: 26.2 KiB, free: 434.3 MiB)
22/07/19 23:41:06 INFO ShuffleBlockFetcherIterator: Getting 1
(2.2 KiB) non-empty blocks including 1 (2.2 KiB) local and 0 (0.0 B) host-local and 0 (0.0 B) push-merged-local and 0 (0.0 B) remote blocks
22/07/19 23:41:06 INFO ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
22/07/19 23:41:06 INFO Executor: Finished task 0.0 in stage 25.0 (TID 24). 3469 bytes result sent to driver
22/07/19 23:41:06 INFO TaskSetManager: Finished task 0.0 in stage 25.0 (TID 24) in 16 ms on 172.30.59.241 (executor driver) (1/1)
22/07/19 23:41:06 INFO TaskSchedulerImpl: Removed TaskSet 25.0, whose tasks have all completed, from pool
22/07/19 23:41:06 INFO DAGScheduler: ResultStage 25 (collectAsMap at RandomForest.scala:1054) finished in 0.037 s
22/07/19 23:41:06 INFO DAGScheduler: Job 16 is finished. Cancelling potential speculative or zombie tasks for this job
22/07/19 23:41:06 INFO TaskSchedulerImpl: Killing all running tasks in stage 25: Stage finished
22/07/19 23:41:06 INFO DAGScheduler: Job 16 finished: collectAsMap at RandomForest.scala:1054, took 0.108499 s
22/07/19 23:41:06 INFO MemoryStore: Block broadcast_42 stored as values in memory (estimated size 2.9 KiB, free 433.7 MiB)
22/07/19 23:41:06 INFO MemoryStore: Block broadcast_42_piece0 stored as bytes in memory (estimated size 764.0 B, free 433.7 MiB)
22/07/19 23:41:06 INFO BlockManagerInfo: Added broadcast_42_piece0 in memory on 172.30.59.241:41055 (size: 764.0 B, free: 434.3 MiB)
22/07/19 23:41:06 INFO SparkContext: Created broadcast 42 from broadcast at RandomForest.scala:293
22/07/19 23:41:06 INFO Instrumentation: [dbfe5743] {"numFeatures":4}
22/07/19 23:41:06 INFO Instrumentation: [dbfe5743] {"numClasses":3}
22/07/19 23:41:06 INFO Instrumentation: [dbfe5743] {"numExamples":113}
22/07/19 23:41:06 INFO Instrumentation: [dbfe5743] {"sumOfWeights":113.0}
22/07/19 23:41:06 INFO MemoryStore: Block broadcast_43 stored as values in memory (estimated size 40.0 B, free 433.7 MiB)
```

```
22/07/19 23:41:06 INFO MemoryStore: Block broadcast_43_piece0 stored as bytes in memory (estimated size 101.0 B, free 433.7 MiB)
22/07/19 23:41:06 INFO BlockManagerInfo: Added broadcast_43_piece0 in memory on 172.30.59.241:41055 (size: 101.0 B, free: 434.3 MiB)
22/07/19 23:41:06 INFO SparkContext: Created broadcast 43 from broadcast at RandomForest.scala:622
22/07/19 23:41:06 INFO SparkContext: Starting job: collectAsMap at RandomForest.scala:663
22/07/19 23:41:06 INFO DAGScheduler: Registering RDD 100 (mapPartitions at RandomForest.scala:644) as input to shuffle 9
22/07/19 23:41:06 INFO DAGScheduler: Got job 17 (collectAsMap at RandomForest.scala:663) with 1 output partitions
22/07/19 23:41:06 INFO DAGScheduler: Final stage: ResultStage 27 (collectAsMap at RandomForest.scala:663)
22/07/19 23:41:06 INFO DAGScheduler: Parents of final stage: List(ShuffleMapStage 26)
22/07/19 23:41:06 INFO DAGScheduler: Missing parents: List(ShuffleMapStage 26)
22/07/19 23:41:06 INFO DAGScheduler: Submitting ShuffleMapStage 26 (MapPartitionsRDD[100] at mapPartitions at RandomForest.scala:644), which has no missing parents
22/07/19 23:41:06 INFO MemoryStore: Block broadcast_44 stored as values in memory (estimated size 70.6 KiB, free 433.6 MiB)
22/07/19 23:41:06 INFO MemoryStore: Block broadcast_44_piece0 stored as bytes in memory (estimated size 28.6 KiB, free 433.6 MiB)
22/07/19 23:41:06 INFO BlockManagerInfo: Added broadcast_44_piece0 in memory on 172.30.59.241:41055 (size: 28.6 KiB, free: 434.3 MiB)
22/07/19 23:41:06 INFO SparkContext: Created broadcast 44 from broadcast at DAGScheduler.scala:1513
22/07/19 23:41:06 INFO DAGScheduler: Submitting 1 missing tasks from ShuffleMapStage 26
 (MapPartitionsRDD[100] at mapPartitions at RandomForest.scala:644) (first 15 tasks are for partitions Vector(0))
22/07/19 23:41:06 INFO TaskSchedulerImpl: Adding task set 26.0 with 1 tasks resource profile 0
22/07/19 23:41:06 INFO TaskSetManager: Starting task 0.0 in stage 26.0 (TID 25)
(172.30.59.241, executor driver, partition 0, PROCESS_LOCAL, 4895 bytes) taskResourceAssignments Map()
22/07/19 23:41:06 INFO Executor: Running task 0.0 in stage 26.0 (TID 25)
22/07/19 23:41:06 INFO FileScanRDD: Reading File path: file:///home/nur_haque/HW6/Iris.csv, range: 0-5107, partition values: [empty row]
22/07/19 23:41:06 INFO MemoryStore: Block rdd_99_0 stored as values in memory (estimated size 11.1 KiB, free 433.6 MiB)
22/07/19 23:41:06 INFO BlockManagerInfo: Added rdd_99_0 in memory on 172.30.59.241:41055 (size: 11.1 KiB, free: 434.3 MiB)
22/07/19 23:41:06 INFO Executor: Finished task 0.0 in stage 26.0 (TID 25). 2093 bytes result sent to driver
22/07/19 23:41:06 INFO TaskSetManager: Finished task 0.0 in stage 26.0 (TID 25) in 51 ms on 172.30.59.241 (executor driver) (1/1)
22/07/19 23:41:06 INFO TaskSchedulerImpl: Removed TaskSet 26.0, whose tasks have all completed, from pool
22/07/19 23:41:06 INFO DAGScheduler: ShuffleMapStage 26 (mapPartitions at RandomForest.scala:644) finished in 0.062 s
22/07/19 23:41:06 INFO DAGScheduler: looking for newly runnable stages
22/07/19 23:41:06 INFO DAGScheduler: running: Set()
22/07/19 23:41:06 INFO DAGScheduler: waiting: Set(ResultStage 27)
22/07/19 23:41:06 INFO DAGScheduler: failed: Set()
22/07/19 23:41:06 INFO DAGScheduler: Submitting ResultStage 27 (MapPartitionsRDD[102] at map at RandomForest.scala:663), which has no missing parents
22/07/19 23:41:06 INFO MemoryStore: Block broadcast_45 stored as values in memory (estimated size 6.9 KiB, free 433.6 MiB)
22/07/19 23:41:06 INFO MemoryStore: Block broadcast_45_piece0 stored as bytes in memory (estimated size 3.7 KiB, free 433.6 MiB)
22/07/19 23:41:06 INFO BlockManagerInfo: Added broadcast_45_piece0 in memory on 172.30.59.241:41055 (size: 3.7 KiB, free: 434.3 MiB)
22/07/19 23:41:06 INFO SparkContext: Created broadcast 45 from broadcast at DAGScheduler.scala:1513
22/07/19 23:41:06 INFO DAGScheduler: Submitting 1 missing tasks from ResultStage 27
(MapPartitionsRDD[102] at map at RandomForest.scala:663) (first 15 tasks are for partitions Vector(0))
22/07/19 23:41:06 INFO TaskSchedulerImpl: Adding task set 27.0 with 1 tasks resource profile 0
22/07/19 23:41:06 INFO TaskSetManager: Starting task 0.0 in stage 27.0
 (TID 26) (172.30.59.241, executor driver, partition 0, NODE_LOCAL, 4271 bytes) taskResourceAssignments Map()
22/07/19 23:41:06 INFO Executor: Running task 0.0 in stage 27.0 (TID 26)
22/07/19 23:41:06 INFO ShuffleBlockFetcherIterator: Getting 1 (2.4 KiB)
non-empty blocks including 1 (2.4 KiB) local and 0 (0.0 B) host-local and 0 (0.0 B) push-merged-local and 0 (0.0 B) remote blocks
22/07/19 23:41:06 INFO ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
22/07/19 23:41:06 INFO Executor: Finished task 0.0 in stage 27.0 (TID 26). 2080 bytes result sent to driver
22/07/19 23:41:06 INFO TaskSetManager: Finished task 0.0 in stage 27.0 (TID 26) in 11 ms on 172.30.59.241 (executor driver) (1/1)
22/07/19 23:41:06 INFO TaskSchedulerImpl: Removed TaskSet 27.0, whose tasks have all completed, from pool
22/07/19 23:41:06 INFO DAGScheduler: ResultStage 27 (collectAsMap at RandomForest.scala:663) finished in 0.018 s
22/07/19 23:41:06 INFO DAGScheduler: Job 17 is finished. Cancelling potential speculative or zombie tasks for this job
22/07/19 23:41:06 INFO TaskSchedulerImpl: Killing all running tasks in stage 27: Stage finished
22/07/19 23:41:06 INFO DAGScheduler: Job 17 finished: collectAsMap at RandomForest.scala:663, took 0.091954 s
22/07/19 23:41:06 INFO TorrentBroadcast: Destroying Broadcast(43) (from destroy at RandomForest.scala:674)
22/07/19 23:41:06 INFO BlockManagerInfo: Removed broadcast_43_piece0 on 172.30.59.241:41055 in memory (size: 101.0 B, free: 434.3 MiB)
22/07/19 23:41:06 INFO MemoryStore: Block broadcast_46 stored as values in memory (estimated size 40.0 B, free 433.6 MiB)
22/07/19 23:41:06 INFO MemoryStore: Block broadcast_46_piece0 stored as bytes in memory (estimated size 101.0 B, free 433.6 MiB)
22/07/19 23:41:06 INFO BlockManagerInfo: Added broadcast_46_piece0 in memory on 172.30.59.241:41055 (size: 101.0 B, free: 434.3 MiB)
22/07/19 23:41:06 INFO SparkContext: Created broadcast 46 from broadcast at RandomForest.scala:622
22/07/19 23:41:06 INFO SparkContext: Starting job: collectAsMap at RandomForest.scala:663
22/07/19 23:41:06 INFO DAGScheduler: Registering RDD 103 (mapPartitions at RandomForest.scala:644) as input to shuffle 10
22/07/19 23:41:06 INFO DAGScheduler: Got job 18 (collectAsMap at RandomForest.scala:663) with 1 output partitions
22/07/19 23:41:06 INFO DAGScheduler: Final stage: ResultStage 29 (collectAsMap at RandomForest.scala:663)
22/07/19 23:41:06 INFO DAGScheduler: Parents of final stage: List(ShuffleMapStage 28)
22/07/19 23:41:06 INFO DAGScheduler: Missing parents: List(ShuffleMapStage 28)
22/07/19 23:41:06 INFO DAGScheduler: Submitting ShuffleMapStage 28 (MapPartitionsRDD[103] at mapPartitions at RandomForest.scala:644), which has no missing parents
22/07/19 23:41:06 INFO MemoryStore: Block broadcast_47 stored as values in memory (estimated size 71.4 KiB, free 433.5 MiB)
22/07/19 23:41:06 INFO MemoryStore: Block broadcast_47_piece0 stored as bytes in memory (estimated size 29.0 KiB, free 433.5 MiB)
22/07/19 23:41:06 INFO BlockManagerInfo: Added broadcast_47_piece0 in memory on 172.30.59.241:41055 (size: 29.0 KiB, free: 434.2 MiB)
22/07/19 23:41:06 INFO SparkContext: Created broadcast 47 from broadcast at DAGScheduler.scala:1513
22/07/19 23:41:06 INFO DAGScheduler: Submitting 1 missing tasks from ShuffleMapStage 28
 (MapPartitionsRDD[103] at mapPartitions at RandomForest.scala:644) (first 15 tasks are for partitions Vector(0))
22/07/19 23:41:06 INFO TaskSchedulerImpl: Adding task set 28.0 with 1 tasks resource profile 0
22/07/19 23:41:06 INFO TaskSetManager: Starting task 0.0 in stage 28.0 (TID 27) (172.30.59.241,
executor driver, partition 0, PROCESS_LOCAL, 4895 bytes) taskResourceAssignments Map()
22/07/19 23:41:06 INFO Executor: Running task 0.0 in stage 28.0 (TID 27)
22/07/19 23:41:06 INFO BlockManager: Found block rdd_99_0 locally
22/07/19 23:41:06 INFO Executor: Finished task 0.0 in stage 28.0 (TID 27). 2093 bytes result sent to driver
22/07/19 23:41:06 INFO TaskSetManager: Finished task 0.0 in stage 28.0 (TID 27) in 17 ms on 172.30.59.241 (executor driver) (1/1)
22/07/19 23:41:06 INFO TaskSchedulerImpl: Removed TaskSet 28.0, whose tasks have all completed, from pool
22/07/19 23:41:06 INFO DAGScheduler: ShuffleMapStage 28 (mapPartitions at RandomForest.scala:644) finished in 0.026 s
22/07/19 23:41:06 INFO DAGScheduler: looking for newly runnable stages
22/07/19 23:41:06 INFO DAGScheduler: running: Set()
22/07/19 23:41:06 INFO DAGScheduler: waiting: Set(ResultStage 29)
22/07/19 23:41:06 INFO DAGScheduler: failed: Set()
22/07/19 23:41:06 INFO DAGScheduler: Submitting ResultStage 29 (MapPartitionsRDD[105] at map at RandomForest.scala:663), which has no missing parents
22/07/19 23:41:06 INFO MemoryStore: Block broadcast_48 stored as values in memory (estimated size 7.4 KiB, free 433.5 MiB)
22/07/19 23:41:06 INFO MemoryStore: Block broadcast_48_piece0 stored as bytes in memory (estimated size 4.0 KiB, free 433.5 MiB)
22/07/19 23:41:06 INFO BlockManagerInfo: Added broadcast_48_piece0 in memory on 172.30.59.241:41055 (size: 4.0 KiB, free: 434.2 MiB)
22/07/19 23:41:06 INFO BlockManagerInfo: Removed broadcast_45_piece0 on 172.30.59.241:41055 in memory (size: 3.7 KiB, free: 434.2 MiB)
22/07/19 23:41:06 INFO SparkContext: Created broadcast 48 from broadcast at DAGScheduler.scala:1513
22/07/19 23:41:06 INFO DAGScheduler: Submitting 1 missing tasks from ResultStage
29 (MapPartitionsRDD[105] at map at RandomForest.scala:663) (first 15 tasks are for partitions Vector(0))
22/07/19 23:41:06 INFO TaskSchedulerImpl: Adding task set 29.0 with 1 tasks resource profile 0
22/07/19 23:41:06 INFO BlockManagerInfo: Removed broadcast_44_piece0 on 172.30.59.241:41055 in memory (size: 28.6 KiB, free: 434.3 MiB)
22/07/19 23:41:06 INFO TaskSetManager: Starting task 0.0 in stage 29.0 (TID 28)
(172.30.59.241, executor driver, partition 0, NODE_LOCAL, 4271 bytes) taskResourceAssignments Map()
22/07/19 23:41:06 INFO Executor: Running task 0.0 in stage 29.0 (TID 28)
22/07/19 23:41:06 INFO BlockManagerInfo: Removed broadcast_41_piece0 on 172.30.59.241:41055 in memory (size: 4.5 KiB, free: 434.3 MiB)
22/07/19 23:41:06 INFO ShuffleBlockFetcherIterator: Getting 1 (2.2 KiB) non-empty
blocks including 1 (2.2 KiB) local and 0 (0.0 B) host-local and 0 (0.0 B) push-merged-local and 0 (0.0 B) remote blocks
22/07/19 23:41:06 INFO ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
22/07/19 23:41:06 INFO BlockManagerInfo: Removed broadcast_40_piece0 on 172.30.59.241:41055 in memory (size: 27.9 KiB, free: 434.3 MiB)
22/07/19 23:41:06 INFO Executor: Finished task 0.0 in stage 29.0 (TID 28). 2080 bytes result sent to driver
22/07/19 23:41:06 INFO TaskSetManager: Finished task 0.0 in stage 29.0 (TID 28) in 20 ms on 172.30.59.241 (executor driver) (1/1)
22/07/19 23:41:06 INFO TaskSchedulerImpl: Removed TaskSet 29.0, whose tasks have all completed, from pool
22/07/19 23:41:06 INFO DAGScheduler: ResultStage 29 (collectAsMap at RandomForest.scala:663) finished in 0.042 s
22/07/19 23:41:06 INFO DAGScheduler: Job 18 is finished. Cancelling potential speculative or zombie tasks for this job
22/07/19 23:41:06 INFO TaskSchedulerImpl: Killing all running tasks in stage 29: Stage finished
22/07/19 23:41:06 INFO DAGScheduler: Job 18 finished: collectAsMap at RandomForest.scala:663, took 0.075203 s
22/07/19 23:41:06 INFO TorrentBroadcast: Destroying Broadcast(46) (from destroy at RandomForest.scala:674)
22/07/19 23:41:06 INFO MemoryStore: Block broadcast_49 stored as values in memory (estimated size 40.0 B, free 433.7 MiB)
22/07/19 23:41:06 INFO BlockManagerInfo: Removed broadcast_46_piece0 on 172.30.59.241:41055 in memory (size: 101.0 B, free: 434.3 MiB)
22/07/19 23:41:06 INFO MemoryStore: Block broadcast_49_piece0 stored as bytes in memory (estimated size 101.0 B, free 433.7 MiB)
22/07/19 23:41:06 INFO BlockManagerInfo: Added broadcast_49_piece0 in memory on 172.30.59.241:41055 (size: 101.0 B, free: 434.3 MiB)
22/07/19 23:41:06 INFO SparkContext: Created broadcast 49 from broadcast at RandomForest.scala:622
22/07/19 23:41:06 INFO SparkContext: Starting job: collectAsMap at RandomForest.scala:663
22/07/19 23:41:06 INFO DAGScheduler: Registering RDD 106 (mapPartitions at RandomForest.scala:644) as input to shuffle 11
22/07/19 23:41:06 INFO DAGScheduler: Got job 19 (collectAsMap at RandomForest.scala:663) with 1 output partitions
22/07/19 23:41:06 INFO DAGScheduler: Final stage: ResultStage 31 (collectAsMap at RandomForest.scala:663)
22/07/19 23:41:06 INFO DAGScheduler: Parents of final stage: List(ShuffleMapStage 30)
22/07/19 23:41:06 INFO DAGScheduler: Missing parents: List(ShuffleMapStage 30)
22/07/19 23:41:06 INFO DAGScheduler: Submitting ShuffleMapStage 30 (MapPartitionsRDD[106] at mapPartitions at RandomForest.scala:644), which has no missing parents
22/07/19 23:41:06 INFO MemoryStore: Block broadcast_50 stored as values in memory (estimated size 71.6 KiB, free 433.6 MiB)
22/07/19 23:41:06 INFO MemoryStore: Block broadcast_50_piece0 stored as bytes in memory (estimated size 29.1 KiB, free 433.6 MiB)
22/07/19 23:41:06 INFO BlockManagerInfo: Added broadcast_50_piece0 in memory on 172.30.59.241:41055 (size: 29.1 KiB, free: 434.3 MiB)
22/07/19 23:41:06 INFO SparkContext: Created broadcast 50 from broadcast at DAGScheduler.scala:1513
22/07/19 23:41:06 INFO DAGScheduler: Submitting 1 missing tasks from ShuffleMapStage
30 (MapPartitionsRDD[106] at mapPartitions at RandomForest.scala:644) (first 15 tasks are for partitions Vector(0))
22/07/19 23:41:06 INFO TaskSchedulerImpl: Adding task set 30.0 with 1 tasks resource profile 0
22/07/19 23:41:06 INFO TaskSetManager: Starting task 0.0 in stage 30.0 (TID 29)
 (172.30.59.241, executor driver, partition 0, PROCESS_LOCAL, 4895 bytes) taskResourceAssignments Map()
22/07/19 23:41:06 INFO Executor: Running task 0.0 in stage 30.0 (TID 29)
22/07/19 23:41:06 INFO BlockManager: Found block rdd_99_0 locally
22/07/19 23:41:06 INFO Executor: Finished task 0.0 in stage 30.0 (TID 29). 2093 bytes result sent to driver
22/07/19 23:41:06 INFO TaskSetManager: Finished task 0.0 in stage 30.0 (TID 29) in 21 ms on 172.30.59.241 (executor driver) (1/1)
```

```
22/07/19 23:41:06 INFO TaskSchedulerImpl: Removed TaskSet 30.0, whose tasks have all completed, from pool
22/07/19 23:41:06 INFO DAGScheduler: ShuffleMapStage 30 (mapPartitions at RandomForest.scala:644) finished in 0.028 s
22/07/19 23:41:06 INFO DAGScheduler: looking for newly runnable stages
22/07/19 23:41:06 INFO DAGScheduler: running: Set()
22/07/19 23:41:06 INFO DAGScheduler: waiting: Set(ResultStage 31)
22/07/19 23:41:06 INFO DAGScheduler: failed: Set()
22/07/19 23:41:06 INFO DAGScheduler: Submitting ResultStage 31 (MapPartitionsRDD[108] at map at RandomForest.scala:663), which has no missing parents
22/07/19 23:41:06 INFO MemoryStore: Block broadcast_51 stored as values in memory (estimated size 7.5 KiB, free 433.6 MiB)
22/07/19 23:41:06 INFO MemoryStore: Block broadcast_51_piece0 stored as bytes in memory (estimated size 4.0 KiB, free 433.6 MiB)
22/07/19 23:41:06 INFO BlockManagerInfo: Added broadcast_51_piece0 in memory on 172.30.59.241:41055 (size: 4.0 KiB, free: 434.3 MiB)
22/07/19 23:41:06 INFO SparkContext: Created broadcast 51 from broadcast at DAGScheduler.scala:1513
22/07/19 23:41:06 INFO DAGScheduler: Submitting 1 missing tasks from ResultStage 31
(MapPartitionsRDD[108] at map at RandomForest.scala:663) (first 15 tasks are for partitions Vector(0))
22/07/19 23:41:06 INFO TaskSchedulerImpl: Adding task set 31.0 with 1 tasks resource profile 0
22/07/19 23:41:06 INFO TaskSetManager: Starting task 0.0 in stage 31.0 (TID 30) (172.30.59.241,
 executor driver, partition 0, NODE_LOCAL, 4271 bytes) taskResourceAssignments Map()
22/07/19 23:41:06 INFO Executor: Running task 0.0 in stage 31.0 (TID 30)
22/07/19 23:41:06 INFO ShuffleBlockFetcherIterator: Getting 1 (2.4 KiB) non-empty blocks
 including 1 (2.4 KiB) local and 0 (0.0 B) host-local and 0 (0.0 B) push-merged-local and 0 (0.0 B) remote blocks
22/07/19 23:41:06 INFO ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
22/07/19 23:41:06 INFO Executor: Finished task 0.0 in stage 31.0 (TID 30). 2289 bytes result sent to driver
22/07/19 23:41:06 INFO TaskSetManager: Finished task 0.0 in stage 31.0 (TID 30) in 13 ms on 172.30.59.241 (executor driver) (1/1)
22/07/19 23:41:06 INFO TaskSchedulerImpl: Removed TaskSet 31.0, whose tasks have all completed, from pool
22/07/19 23:41:06 INFO DAGScheduler: ResultStage 31 (collectAsMap at RandomForest.scala:663) finished in 0.027 s
22/07/19 23:41:06 INFO DAGScheduler: Job 19 is finished. Cancelling potential speculative or zombie tasks for this job
22/07/19 23:41:06 INFO TaskSchedulerImpl: Killing all running tasks in stage 31: Stage finished
22/07/19 23:41:06 INFO DAGScheduler: Job 19 finished: collectAsMap at RandomForest.scala:663, took 0.067028 s
22/07/19 23:41:06 INFO TorrentBroadcast: Destroying Broadcast(49) (from destroy at RandomForest.scala:674)
22/07/19 23:41:06 INFO BlockManagerInfo: Removed broadcast_49_piece0 on 172.30.59.241:41055 in memory (size: 101.0 B, free: 434.3 MiB)
22/07/19 23:41:06 INFO MemoryStore: Block broadcast_52 stored as values in memory (estimated size 40.0 B, free 433.6 MiB)
22/07/19 23:41:06 INFO MemoryStore: Block broadcast_52_piece0 stored as bytes in memory (estimated size 101.0 B, free 433.6 MiB)
22/07/19 23:41:06 INFO BlockManagerInfo: Added broadcast_52_piece0 in memory on 172.30.59.241:41055 (size: 101.0 B, free: 434.3 MiB)
22/07/19 23:41:06 INFO SparkContext: Created broadcast 52 from broadcast at RandomForest.scala:622
22/07/19 23:41:06 INFO SparkContext: Starting job: collectAsMap at RandomForest.scala:663
22/07/19 23:41:06 INFO DAGScheduler: Registering RDD 109 (mapPartitions at RandomForest.scala:644) as input to shuffle 12
22/07/19 23:41:06 INFO DAGScheduler: Got job 20 (collectAsMap at RandomForest.scala:663) with 1 output partitions
22/07/19 23:41:06 INFO DAGScheduler: Final stage: ResultStage 33 (collectAsMap at RandomForest.scala:663)
22/07/19 23:41:06 INFO DAGScheduler: Parents of final stage: List(ShuffleMapStage 32)
22/07/19 23:41:06 INFO DAGScheduler: Missing parents: List(ShuffleMapStage 32)
22/07/19 23:41:06 INFO DAGScheduler: Submitting ShuffleMapStage 32 (MapPartitionsRDD[109] at mapPartitions at RandomForest.scala:644), which has no missing parents
22/07/19 23:41:06 INFO MemoryStore: Block broadcast_53 stored as values in memory (estimated size 72.2 KiB, free 433.5 MiB)
22/07/19 23:41:06 INFO MemoryStore: Block broadcast_53_piece0 stored as bytes in memory (estimated size 29.3 KiB, free 433.5 MiB)
22/07/19 23:41:06 INFO BlockManagerInfo: Added broadcast_53_piece0 in memory on 172.30.59.241:41055 (size: 29.3 KiB, free: 434.2 MiB)
22/07/19 23:41:06 INFO SparkContext: Created broadcast 53 from broadcast at DAGScheduler.scala:1513
22/07/19 23:41:06 INFO DAGScheduler: Submitting 1 missing tasks from ShuffleMapStage 32
(MapPartitionsRDD[109] at mapPartitions at RandomForest.scala:644) (first 15 tasks are for partitions Vector(0))
22/07/19 23:41:06 INFO TaskSchedulerImpl: Adding task set 32.0 with 1 tasks resource profile 0
22/07/19 23:41:06 INFO TaskSetManager: Starting task 0.0 in stage 32.0 (TID 31)
(172.30.59.241, executor driver, partition 0, PROCESS_LOCAL, 4895 bytes) taskResourceAssignments Map()
22/07/19 23:41:06 INFO Executor: Running task 0.0 in stage 32.0 (TID 31)
22/07/19 23:41:06 INFO BlockManager: Found block rdd_99_0 locally
22/07/19 23:41:06 INFO Executor: Finished task 0.0 in stage 32.0 (TID 31). 2093 bytes result sent to driver
22/07/19 23:41:06 INFO TaskSetManager: Finished task 0.0 in stage 32.0 (TID 31) in 18 ms on 172.30.59.241 (executor driver) (1/1)
22/07/19 23:41:06 INFO TaskSchedulerImpl: Removed TaskSet 32.0, whose tasks have all completed, from pool
22/07/19 23:41:06 INFO DAGScheduler: ShuffleMapStage 32 (mapPartitions at RandomForest.scala:644) finished in 0.027 s
22/07/19 23:41:06 INFO DAGScheduler: looking for newly runnable stages
22/07/19 23:41:06 INFO DAGScheduler: running: Set()
22/07/19 23:41:06 INFO DAGScheduler: waiting: Set(ResultStage 33)
22/07/19 23:41:06 INFO DAGScheduler: failed: Set()
22/07/19 23:41:06 INFO DAGScheduler: Submitting ResultStage 33 (MapPartitionsRDD[111]
 at map at RandomForest.scala:663), which has no missing parents
22/07/19 23:41:06 INFO MemoryStore: Block broadcast_54 stored as values in memory (estimated size 7.5 KiB, free 433.5 MiB)
22/07/19 23:41:06 INFO MemoryStore: Block broadcast_54_piece0 stored as bytes in memory (estimated size 4.0 KiB, free 433.5 MiB)
22/07/19 23:41:06 INFO BlockManagerInfo: Added broadcast_54_piece0 in memory on 172.30.59.241:41055 (size: 4.0 KiB, free: 434.2 MiB)
22/07/19 23:41:06 INFO SparkContext: Created broadcast 54 from broadcast at DAGScheduler.scala:1513
22/07/19 23:41:06 INFO DAGScheduler: Submitting 1 missing tasks from ResultStage 33
 (MapPartitionsRDD[111] at map at RandomForest.scala:663) (first 15 tasks are for partitions Vector(0))
22/07/19 23:41:06 INFO TaskSchedulerImpl: Adding task set 33.0 with 1 tasks resource profile 0
22/07/19 23:41:06 INFO TaskSetManager: Starting task 0.0 in stage 33.0 (TID 32) (172.30.59.241,
executor driver, partition 0, NODE_LOCAL, 4271 bytes) taskResourceAssignments Map()
22/07/19 23:41:06 INFO Executor: Running task 0.0 in stage 33.0 (TID 32)
22/07/19 23:41:06 INFO ShuffleBlockFetcherIterator: Getting 1 (2.2 KiB) non-empty blocks
including 1 (2.2 KiB) local and 0 (0.0 B) host-local and 0 (0.0 B) push-merged-local and 0 (0.0 B) remote blocks
22/07/19 23:41:06 INFO ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
22/07/19 23:41:06 INFO Executor: Finished task 0.0 in stage 33.0 (TID 32). 2496 bytes result sent to driver
22/07/19 23:41:06 INFO TaskSetManager: Finished task 0.0 in stage 33.0 (TID 32) in 9 ms on 172.30.59.241 (executor driver) (1/1)
22/07/19 23:41:06 INFO TaskSchedulerImpl: Removed TaskSet 33.0, whose tasks have all completed, from pool
22/07/19 23:41:06 INFO DAGScheduler: ResultStage 33 (collectAsMap at RandomForest.scala:663) finished in 0.018 s
22/07/19 23:41:06 INFO DAGScheduler: Job 20 is finished. Cancelling potential speculative or zombie tasks for this job
22/07/19 23:41:06 INFO TaskSchedulerImpl: Killing all running tasks in stage 33: Stage finished
22/07/19 23:41:06 INFO DAGScheduler: Job 20 finished: collectAsMap at RandomForest.scala:663, took 0.054400 s
22/07/19 23:41:06 INFO TorrentBroadcast: Destroying Broadcast(52) (from destroy at RandomForest.scala:674)
22/07/19 23:41:06 INFO RandomForest: Internal timing for DecisionTree:
22/07/19 23:41:06 INFO RandomForest:   init: 2.0E-4
  total: 0.5292663
  findBestSplits: 0.5246646
  chooseSplits: 0.5230664
22/07/19 23:41:06 INFO BlockManagerInfo: Removed broadcast_52_piece0 on 172.30.59.241:41055 in memory (size: 101.0 B, free: 434.2 MiB)
22/07/19 23:41:06 INFO MapPartitionsRDD: Removing RDD 99 from persistence list
22/07/19 23:41:06 INFO BlockManager: Removing RDD 99
22/07/19 23:41:06 INFO TorrentBroadcast: Destroying Broadcast(42) (from destroy at RandomForest.scala:305)
22/07/19 23:41:06 INFO Instrumentation: [dbfe5743] training finished
22/07/19 23:41:06 INFO BlockManagerInfo: Removed broadcast_42_piece0 on 172.30.59.241:41055 in memory (size: 764.0 B, free: 434.2 MiB)
22/07/19 23:41:07 INFO FileSourceStrategy: Pushed Filters:
22/07/19 23:41:07 INFO FileSourceStrategy: Post-Scan Filters:
22/07/19 23:41:07 INFO FileSourceStrategy: Output Data Schema: struct<SepalLengthCm: string, SepalWidthCm:
 string, PetalLengthCm: string, PetalWidthCm: string, Species: string ... 3 more fields>
22/07/19 23:41:07 INFO MemoryStore: Block broadcast_55 stored as values in memory (estimated size 173.6 KiB, free 433.3 MiB)
22/07/19 23:41:07 INFO MemoryStore: Block broadcast_55_piece0 stored as bytes in memory (estimated size 24.7 KiB, free 433.3 MiB)
22/07/19 23:41:07 INFO BlockManagerInfo: Added broadcast_55_piece0 in memory on 172.30.59.241:41055 (size: 24.7 KiB, free: 434.2 MiB)
22/07/19 23:41:07 INFO SparkContext: Created broadcast 55 from rdd at MulticlassClassificationEvaluator.scala:197
22/07/19 23:41:07 INFO FileSourceScanExec: Planning scan with bin packing, max size: 4194304 bytes, open cost is considered as scanning 4194304 bytes.
22/07/19 23:41:07 INFO SparkContext: Starting job: collectAsMap at MulticlassMetrics.scala:61
22/07/19 23:41:07 INFO DAGScheduler: Registering RDD 119 (map at MulticlassMetrics.scala:52) as input to shuffle 13
22/07/19 23:41:07 INFO DAGScheduler: Got job 21 (collectAsMap at MulticlassMetrics.scala:61) with 1 output partitions
22/07/19 23:41:07 INFO DAGScheduler: Final stage: ResultStage 35 (collectAsMap at MulticlassMetrics.scala:61)
22/07/19 23:41:07 INFO DAGScheduler: Parents of final stage: List(ShuffleMapStage 34)
22/07/19 23:41:07 INFO DAGScheduler: Missing parents: List(ShuffleMapStage 34)
22/07/19 23:41:07 INFO DAGScheduler: Submitting ShuffleMapStage 34 (MapPartitionsRDD[119] at map at MulticlassMetrics.scala:52), which has no missing parents
22/07/19 23:41:07 INFO MemoryStore: Block broadcast_56 stored as values in memory (estimated size 90.7 KiB, free 433.2 MiB)
22/07/19 23:41:07 INFO MemoryStore: Block broadcast_56_piece0 stored as bytes in memory (estimated size 35.5 KiB, free 433.2 MiB)
22/07/19 23:41:07 INFO BlockManagerInfo: Added broadcast_56_piece0 in memory on 172.30.59.241:41055 (size: 35.5 KiB, free: 434.2 MiB)
22/07/19 23:41:07 INFO SparkContext: Created broadcast 56 from broadcast at DAGScheduler.scala:1513
22/07/19 23:41:07 INFO DAGScheduler: Submitting 1 missing tasks from ShuffleMapStage 34 (MapPartitionsRDD[119]
 at map at MulticlassMetrics.scala:52) (first 15 tasks are for partitions Vector(0))
22/07/19 23:41:07 INFO TaskSchedulerImpl: Adding task set 34.0 with 1 tasks resource profile 0
22/07/19 23:41:07 INFO TaskSetManager: Starting task 0.0 in stage 34.0 (TID 33) (172.30.59.241,
 executor driver, partition 0, PROCESS_LOCAL, 4895 bytes) taskResourceAssignments Map()
22/07/19 23:41:07 INFO Executor: Running task 0.0 in stage 34.0 (TID 33)
22/07/19 23:41:07 INFO BlockManagerInfo: Removed broadcast_37_piece0 on 172.30.59.241:41055 in memory (size: 24.7 KiB, free: 434.2 MiB)
22/07/19 23:41:07 INFO BlockManagerInfo: Removed broadcast_53_piece0 on 172.30.59.241:41055 in memory (size: 29.3 KiB, free: 434.2 MiB)
22/07/19 23:41:07 INFO BlockManagerInfo: Removed broadcast_36_piece0 on 172.30.59.241:41055 in memory (size: 24.7 KiB, free: 434.2 MiB)
22/07/19 23:41:07 INFO BlockManagerInfo: Removed broadcast_47_piece0 on 172.30.59.241:41055 in memory (size: 29.0 KiB, free: 434.3 MiB)
22/07/19 23:41:07 INFO BlockManagerInfo: Removed broadcast_48_piece0 on 172.30.59.241:41055 in memory (size: 4.0 KiB, free: 434.3 MiB)
22/07/19 23:41:07 INFO BlockManagerInfo: Removed broadcast_54_piece0 on 172.30.59.241:41055 in memory (size: 4.0 KiB, free: 434.3 MiB)
22/07/19 23:41:07 INFO BlockManagerInfo: Removed broadcast_50_piece0 on 172.30.59.241:41055 in memory (size: 29.1 KiB, free: 434.3 MiB)
22/07/19 23:41:07 INFO BlockManagerInfo: Removed broadcast_51_piece0 on 172.30.59.241:41055 in memory (size: 4.0 KiB, free: 434.3 MiB)
22/07/19 23:41:07 INFO BlockManager: Removing RDD 99
22/07/19 23:41:07 INFO FileScanRDD: Reading File path: file:///home/nur_haque/HW6/Iris.csv, range: 0-5107, partition values: [empty row]
22/07/19 23:41:07 INFO Executor: Finished task 0.0 in stage 34.0 (TID 33). 2093 bytes result sent to driver
22/07/19 23:41:07 INFO TaskSetManager: Finished task 0.0 in stage 34.0 (TID 33) in 117 ms on 172.30.59.241 (executor driver) (1/1)
22/07/19 23:41:07 INFO TaskSchedulerImpl: Removed TaskSet 34.0, whose tasks have all completed, from pool
22/07/19 23:41:07 INFO DAGScheduler: ShuffleMapStage 34 (map at MulticlassMetrics.scala:52) finished in 0.152 s
22/07/19 23:41:07 INFO DAGScheduler: looking for newly runnable stages
22/07/19 23:41:07 INFO DAGScheduler: running: Set()
```

```
22/07/19 23:41:07 INFO DAGScheduler: waiting: Set(ResultStage 35)
22/07/19 23:41:07 INFO DAGScheduler: failed: Set()
22/07/19 23:41:07 INFO DAGScheduler: Submitting ResultStage 35 (ShuffledRDD[120] at reduceByKey at MulticlassMetrics.scala:61), which has no missing parents
22/07/19 23:41:07 INFO MemoryStore: Block broadcast_57 stored as values in memory (estimated size 4.7 KiB, free 433.9 MiB)
22/07/19 23:41:07 INFO MemoryStore: Block broadcast_57_piece0 stored as bytes in memory (estimated size 2.8 KiB, free 433.9 MiB)
22/07/19 23:41:07 INFO BlockManagerInfo: Added broadcast_57_piece0 in memory on 172.30.59.241:41055 (size: 2.8 KiB, free: 434.3 MiB)
22/07/19 23:41:07 INFO SparkContext: Created broadcast 57 from broadcast at DAGScheduler.scala:1513
22/07/19 23:41:07 INFO DAGScheduler: Submitting 1 missing tasks from ResultStage 35
 (ShuffledRDD[120] at reduceByKey at MulticlassMetrics.scala:61) (first 15 tasks are for partitions Vector(0))
22/07/19 23:41:07 INFO TaskSchedulerImpl: Adding task set 35.0 with 1 tasks resource profile 0
22/07/19 23:41:07 INFO TaskSetManager: Starting task 0.0 in stage 35.0 (TID 34)
(172.30.59.241, executor driver, partition 0, NODE_LOCAL, 4271 bytes) taskResourceAssignments Map()
22/07/19 23:41:07 INFO Executor: Running task 0.0 in stage 35.0 (TID 34)
22/07/19 23:41:07 INFO ShuffleBlockFetcherIterator: Getting 1 (334.0 B)
non-empty blocks including 1 (334.0 B) local and 0 (0.0 B) host-local and 0 (0.0 B) push-merged-local and 0 (0.0 B) remote blocks
22/07/19 23:41:07 INFO ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
22/07/19 23:41:07 INFO Executor: Finished task 0.0 in stage 35.0 (TID 34). 1678 bytes result sent to driver
22/07/19 23:41:07 INFO TaskSetManager: Finished task 0.0 in stage 35.0 (TID 34) in 11 ms on 172.30.59.241 (executor driver) (1/1)
22/07/19 23:41:07 INFO TaskSchedulerImpl: Removed TaskSet 35.0, whose tasks have all completed, from pool
22/07/19 23:41:07 INFO DAGScheduler: ResultStage 35 (collectAsMap at MulticlassMetrics.scala:61) finished in 0.018 s
22/07/19 23:41:07 INFO DAGScheduler: Job 21 is finished. Cancelling potential speculative or zombie tasks for this job
22/07/19 23:41:07 INFO TaskSchedulerImpl: Killing all running tasks in stage 35: Stage finished
22/07/19 23:41:07 INFO DAGScheduler: Job 21 finished: collectAsMap at MulticlassMetrics.scala:61, took 0.178383 s
Test Accuracy of DecisionTreeClassifier = 0.9459459459459459
22/07/19 23:41:07 INFO SparkContext: Invoking stop() from shutdown hook
22/07/19 23:41:07 INFO SparkUI: Stopped Spark web UI at http://172.30.59.241:4040
22/07/19 23:41:07 INFO MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!
22/07/19 23:41:07 INFO MemoryStore: MemoryStore cleared
22/07/19 23:41:07 INFO BlockManager: BlockManager stopped
22/07/19 23:41:07 INFO BlockManagerMaster: BlockManagerMaster stopped
22/07/19 23:41:07 INFO OutputCommitCoordinator$OutputCommitCoordinatorEndpoint: OutputCommitCoordinator stopped!
22/07/19 23:41:07 INFO SparkContext: Successfully stopped SparkContext
22/07/19 23:41:07 INFO ShutdownHookManager: Shutdown hook called
22/07/19 23:41:07 INFO ShutdownHookManager: Deleting directory /tmp/spark-934e49c5-5591-4b1f-8f33-19a33e0e810e
22/07/19 23:41:07 INFO ShutdownHookManager: Deleting directory /tmp/spark-35312ccd-940f-4511-82eb-7c6331660955/pyspark-a04c0556-5bdb-4217-ad6b-95485494e414
22/07/19 23:41:07 INFO ShutdownHookManager: Deleting directory /tmp/spark-35312ccd-940f-4511-82eb-7c6331660955
```