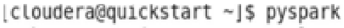
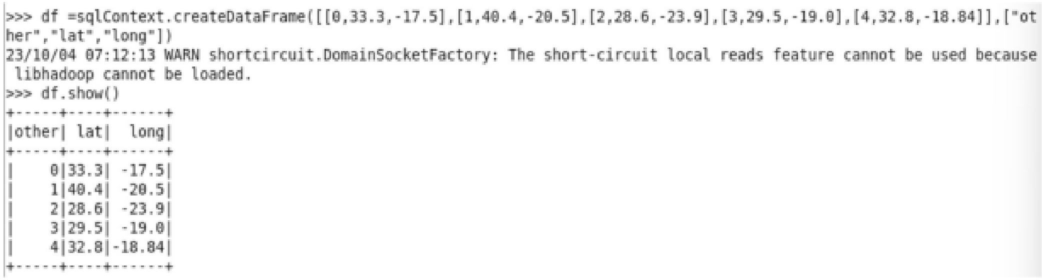
Vivekanand Education Society's
Institute of Technology
Approved by AICTE & Affiliated to University of Mumbai

Artificial Intelligence and Data Science Department

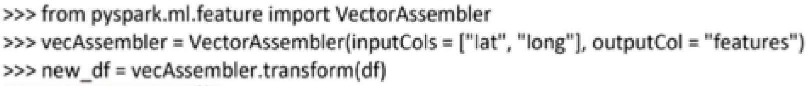
BDA/Odd Sem 2023-23/Experiment 6



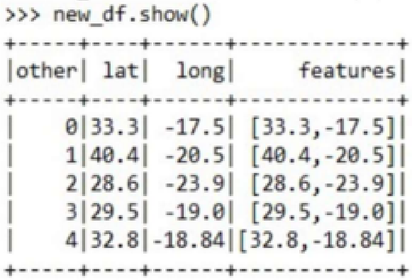
```
[cloudera@quickstart ~]$ pyspark
```



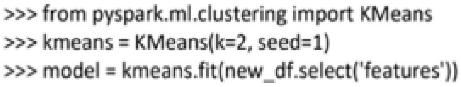
```
>>> df =sqlContext.createDataFrame([[0,33.3,-17.5],[1,40.4,-20.5],[2,28.6,-23.9],[3,29.5,-19.0],[4,32.8,-18.84]],["ot
her","lat","long"])
23/10/04 07:12:13 WARN shortcircuit.DomainSocketFactory: The short-circuit local reads feature cannot be used because
 libhadoop cannot be loaded.
>>> df.show()
+-----+----+------+
|other| lat|  long|
+-----+----+------+
|    0|33.3| -17.5|
|    1|40.4| -20.5|
|    2|28.6| -23.9|
|    3|29.5| -19.0|
|    4|32.8|-18.84|
+-----+----+------+
```



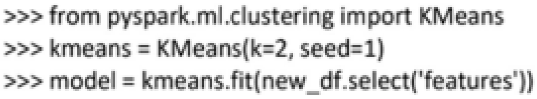
```
>>> from pyspark.ml.feature import VectorAssembler
>>> vecAssembler = VectorAssembler(inputCols = ["lat", "long"], outputCol = "features")
>>> new_df = vecAssembler.transform(df)
```



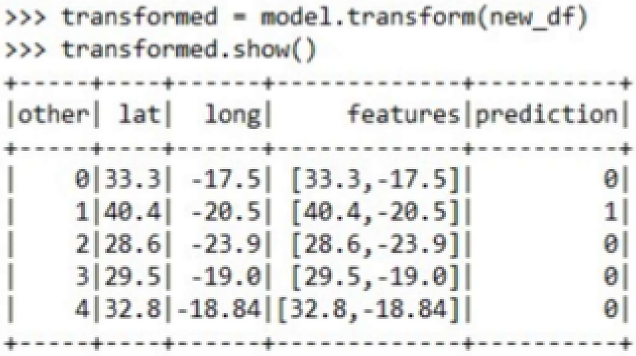
```
>>> new_df.show()
+-----+----+------+-------------+
|other| lat|  long|     features|
+-----+----+------+-------------+
|    0|33.3| -17.5| [33.3,-17.5]|
|    1|40.4| -20.5| [40.4,-20.5]|
|    2|28.6| -23.9| [28.6,-23.9]|
|    3|29.5| -19.0| [29.5,-19.0]|
|    4|32.8|-18.84|[32.8,-18.84]|
+-----+----+------+-------------+
```



```
>>> from pyspark.ml.clustering import KMeans
>>> kmeans = KMeans(k=2, seed=1)
>>> model = kmeans.fit(new_df.select('features'))
```



```
>>> from pyspark.ml.clustering import KMeans
>>> kmeans = KMeans(k=2, seed=1)
>>> model = kmeans.fit(new_df.select('features'))
```



```
>>> transformed = model.transform(new_df)
>>> transformed.show()
+-----+----+------+-------------+----------+
|other| lat|  long|     features|prediction|
+-----+----+------+-------------+----------+
|    0|33.3| -17.5| [33.3,-17.5]|         0|
|    1|40.4| -20.5| [40.4,-20.5]|         1|
|    2|28.6| -23.9| [28.6,-23.9]|         0|
|    3|29.5| -19.0| [29.5,-19.0]|         0|
|    4|32.8|-18.84|[32.8,-18.84]|         0|
+-----+----+------+-------------+----------+
```

Results and Discussions: