Algorithm for K-NN in Hadoop:

1. $hadoop jar <path-to-jar-file> class <path--to-training> <path-to-output> input1/testdata.
2. Array arr <- testdatapoints.
3. Start the job.
4. Map training points.
5. Calculate euclidean\_distance(test\_data\_point,training\_data\_point).
6. Write key-value pair: <index\_test\_point,(euclidean\_distance,class\_training\_point)>.
7. Reduce the key-value pairs based on key.
8. Heap h <- (euclidean\_distance,class\_training\_point).
9. Calculate max\_class\_frequency(h) in first k points.
10. Compare chosen class with class in test datapoint.
11. Write value to ouptut.