Algorithm for k-NN in Spark:

1. Call Spark application with <path-to-training> <path-to-test> <path-to-output>
2. RDD a <- Training\_points
3. RDD b <- Test\_points
4. Form a PairRDD with cross product of a and b.
5. Calculate euclidean\_distance(a\_point,b\_point).
6. Write an RDD of key-value pairs as <key,<euclidean\_distance,training\_class>>.
7. Write an RDD that groups on key.
8. Select classification based on majority.
9. Write result to output file.