1. **Putting dataset into hdfs:** bin/hdfs dfs -put ~/20newsdataall /user/20newsdata
2. **Going to /usr/local/mahout:** cd $MAHOUT\_HOME
3. **Creating sequence file:** bin/mahout seqdirectory -i /user/20newsdata -o /user/20newsdataseq
4. **Creating normalised vectors from sequence files:** bin/mahout seq2sparse -i /user/20newsdataseq/part-m-00000 -o /user/20newsvector -lnorm -nv -wt tfidf
5. **Splitting dataset into training and test set:** bin/mahout split -i /user/20newsvector/tfidf-vectors--trainingOutput /user/20newsdatatrain --testOutput /user/20newsdatatest --randomSelectionPct 40 --overwrite --sequenceFiles -xm sequential
6. **Creating a Naïve Bayes Classification model from training set:** bin/mahout trainnb -i /user/20newsdatatrain -o /user/model -li /user/labelindex -ow -c
7. **Testing and evaluating our model over new Test set:** bin/mahout testnb -i /user/20newsdatatest -m /user/model -l /user/labelindex -o /user/results -ow

**EXTRAS**

**Deleting files from hdfs:** hadoop fs -rmr /user/20newsdatatest

**Mahout UI:** localhost:50070

**Monitoring all daemons:** localhost:8088