DS222 Assignment 1

Bharat Kumar Padi [1]

1. Accuracies(in percentage)

Naive Bayes Local:

Test data: 94.12%

Dev data: 94.38%

Train data: 95.68%

Naive Bayes Map Reduce:

Test data: 79.10%

Dev data: 79.86%

Train data: 84.52%

2. Number of parameters:

number of parameters: 1075832

3. Timings: in secs

Naive Bayes Local:

Train time: 631.8s

Test data: 42s

Naive Bayes Map Reduce:

Train time: 140s

Test data: 451s

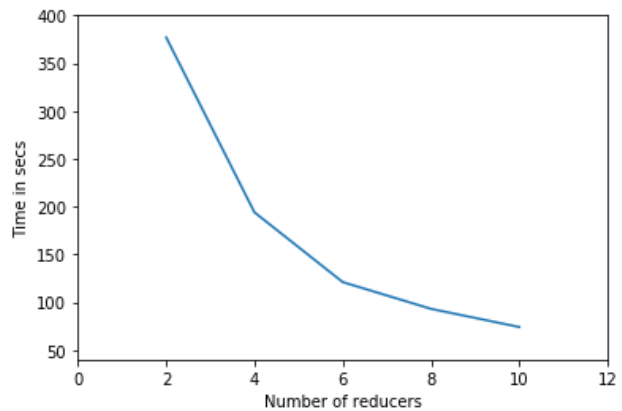### 3.1. Reducer wall clock time plot:

Figure 1. Time taken for reduce job vs Number of reducers

Reduction in the amount of time taken with the increase in the number of reducers is almost exponential, the time taken almost reduced by half initially when the number of reducers increased from 2 to 4 and from 4 to 6, after the because of the overheads involved the reduction in time taken is not significant as the dataset is not too large.

*Equal contribution [1]IISc, Bengaluru, India. Correspondence to: Bharat Kumar Padi <kishna.bharat@com>.