

Machine Learning with Large Datasets

04-04-00-10-42-16-1-13300

Assignment 1

September 23, 2017

1. **Results:** *Training Time For InMemory Naive Bayes :25 seconds*

Table 1: Accuracies of In Memory Naive Bayes in %

Train	Test	Development
96.8	95.86	95.99

Table 2: Accuracies of Hadoop Implementation of Naive Bayes in %

Train	Test	Development
95.96	72.12	76.56

Table 3: Train and Test Timings i seconds

Reducers	Test on Train Data	Test on Test Data	Test on Development Data
In Memory	473	102	176
1	980	120	202
2	720	90	160
4	550	75	109
6	390	49	70
8	220	30	49
10	120	30	40

2. **Number of Parameters in the model were:**

- (a) Words = 448664
- (b) Labels = 50
- (c) Total Number of Parameters = $448664 * 50 = 22433200$