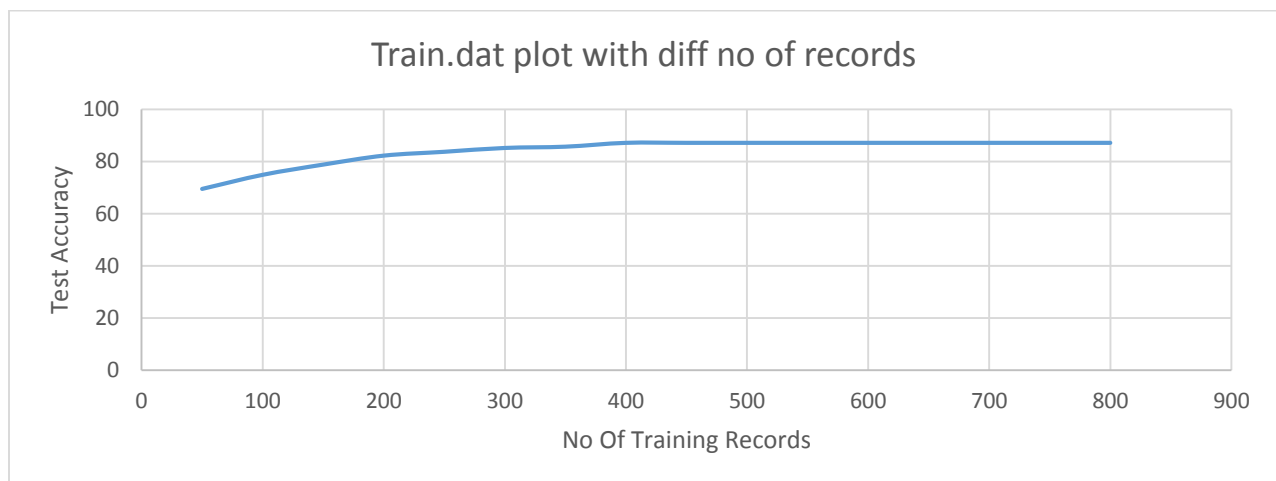


### Part – II Question D

The following analysis is done for the **train.dat** where the decision tree was formed with increasing the train data in the regular interval of 50 records and the accuracy is noted.

No fo Training records	Test Accuracy
50	69.45
100	74.87
150	78.81
200	82.26
250	83.74
300	85.22
350	85.71
400	87.19
450	87.19
500	87.19
550	87.19
600	87.19
650	87.19
700	87.19
750	87.19
800	87.19



The graph depicts that the test accuracy increases as the tree is fed with more training records. The curve stabilized after reaching 400 data inputs and produced the same accuracy even after fed 800 data inputs. The usual property of the learning curve is that it increases accuracy till

some point and later if the tree is over fitted by supplying more train data the accuracy should decrease. The curve plotted above deviates slightly from the expected behavior as we expect the test accuracy to drop after some point. This is may be due to the fact our train data has only 6 attributes and since we are giving the more and more training examples the examples which we feed after 400 train records may be duplicates to the already existing records and so the accuracy didn't drop even when we increase the train data size.