# Data Mining Assignment 3

The assignment implements active learning algorithm on two different datasets, MindReading and MMI using two different sampling based approaches: 1) Random sampling and 2) Uncertainty-based sampling.
Let's talk about the MindReading dataset first. As it can be seen in the given graph of accuracy against the iteration number, the random sampling and uncertainty based sampling both work pretty similar in terms of accuracy. For both techniques, the accuracy starts with value about 0.44 for iteration 1 and then grows till about 0.70 for uncertainty-based sampling and about 0.69 for random sampling in an almost linear fashion.
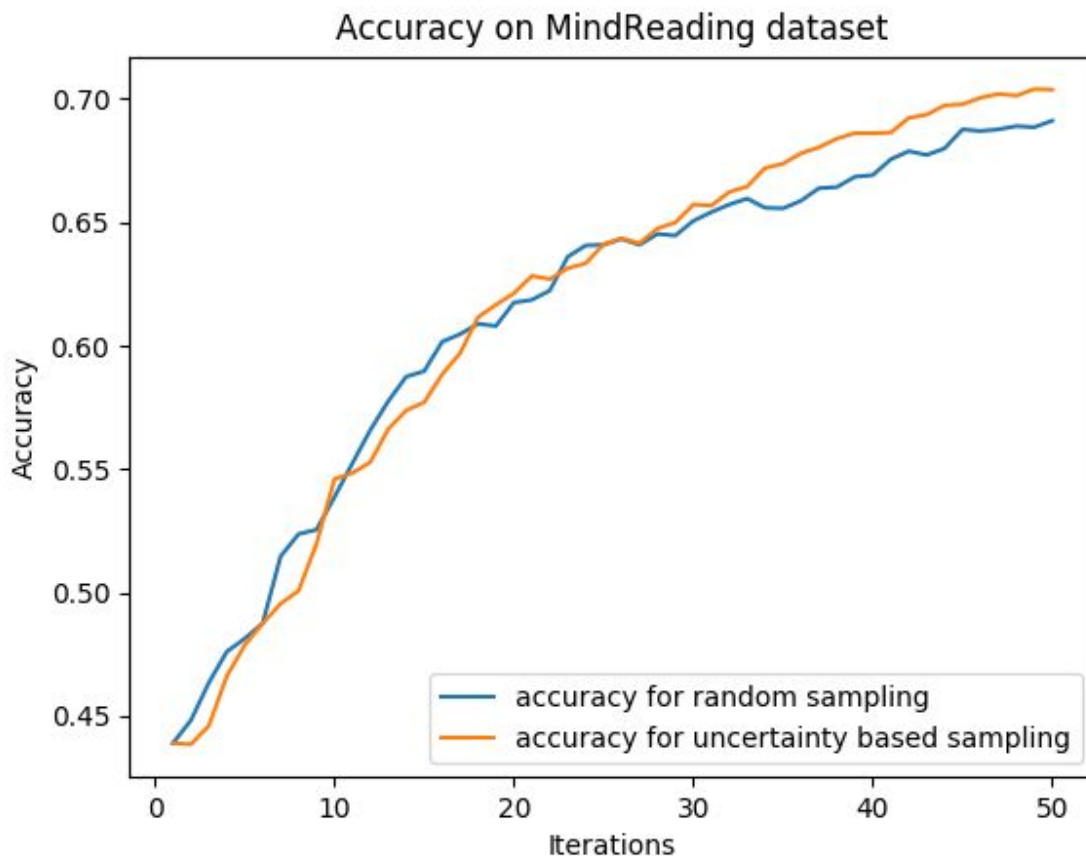
Now, let's talk about the MMI dataset. We see a very similar trend considering the shape of the graph. The random sampling and uncertainty based sampling work pretty similar. However, the accuracy value starts with about 0.57 at iteration 1 and goes till about 0.94 and 0.92 for uncertainty based and random sampling respectively.

Accuracy on MMI dataset
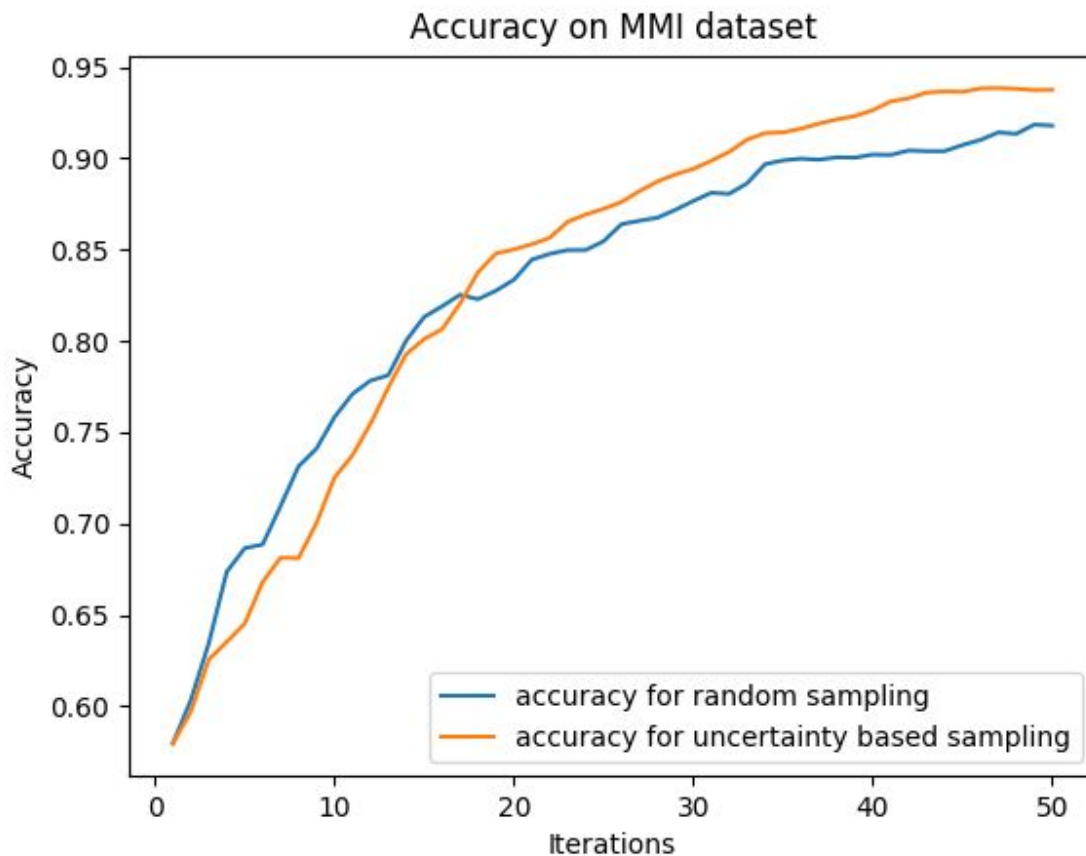
These results show that the logistic regression model relatively better for the MMI dataset than MindReading dataset, although there is not much difference in the accuracy gain for random and uncertainty based sampling approaches within the models.

I used sklearn library in Python for building logistic regression model. For training the model, I used some specific parameters. These parameters include 'newton-cg' as the solver parameter and 'multinomial' as the multi_class parameter to the model. I am rather unfamiliar with the underlying implementations of these parameters. However, given my limited research, I chose 'newton-cg' and 'multinomial' because these work well for multiclass classification. The accuracy value is depending on these parameters, hence it might give a different accuracy model given some other parameters. I have not studied any other ones.