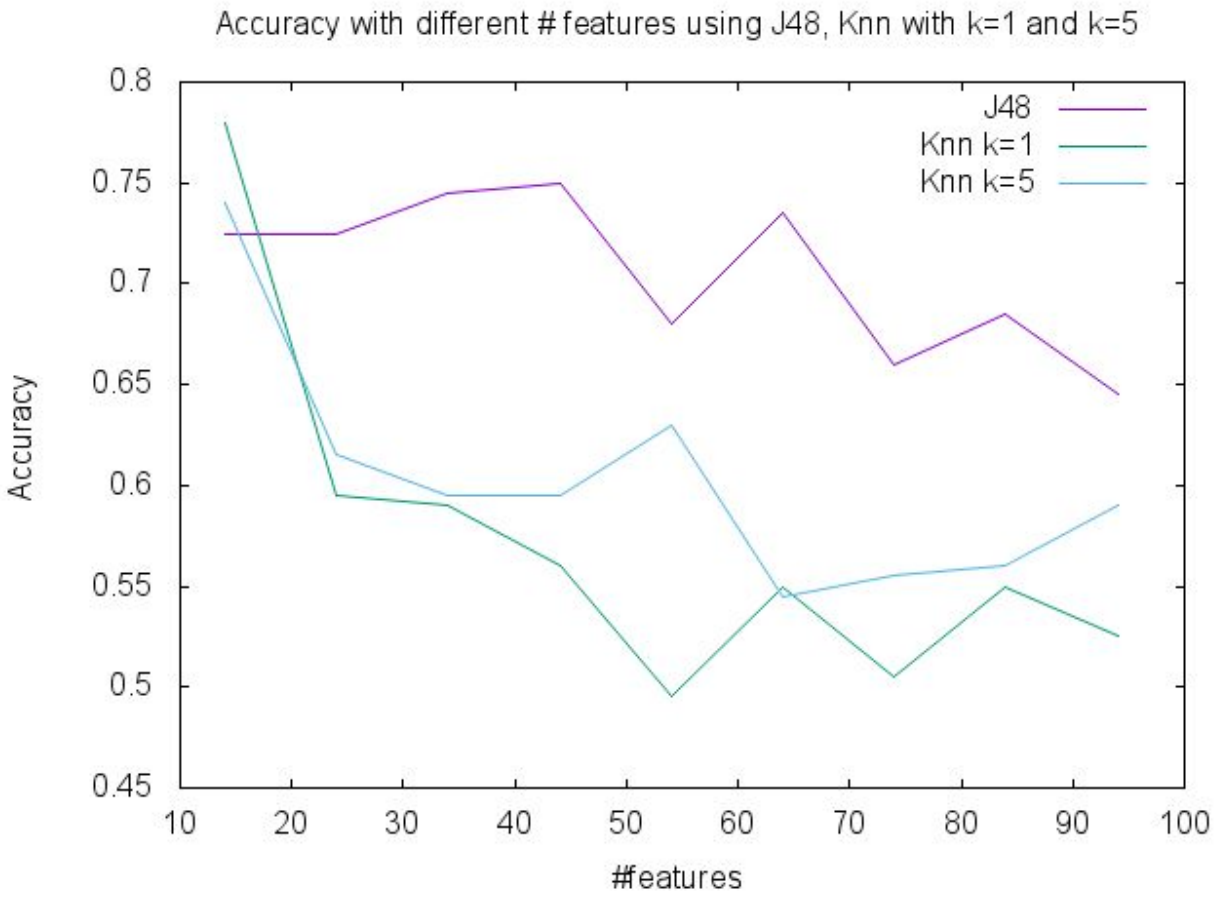
Report for Project 1

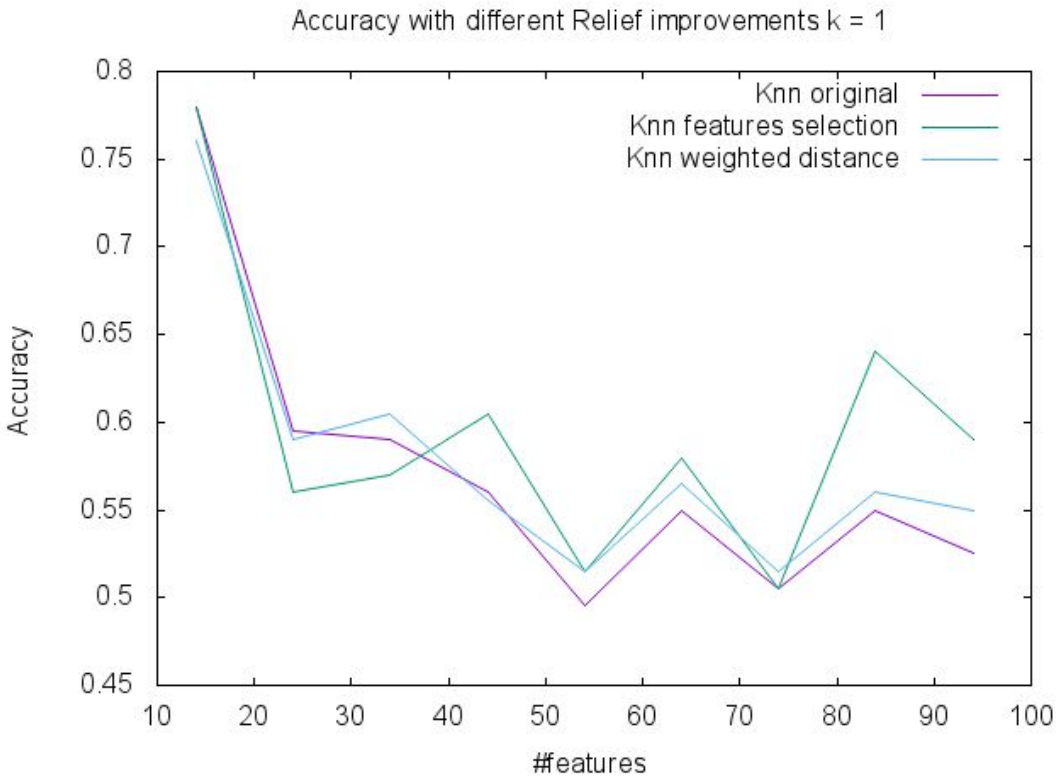
Zhaokun Xue

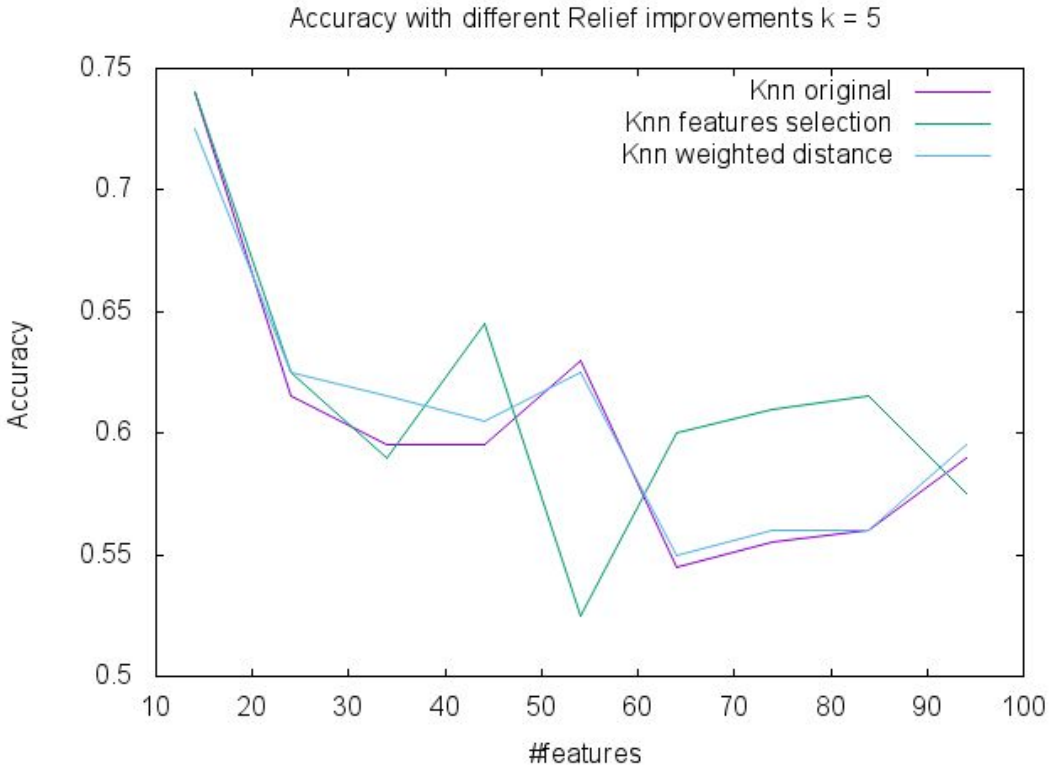
- **Analysis for 3.1 and 3.3**



Based on the output of J48, and KNN method with k = 1 nd k =5, we can tell that J48 method is not so sensitive on the number of features in the data. But KNN method is more sensitive on the number of features in the data. With more features, the accuracy of KNN decreases. For larger k value, the accuracy of KNN is improved.

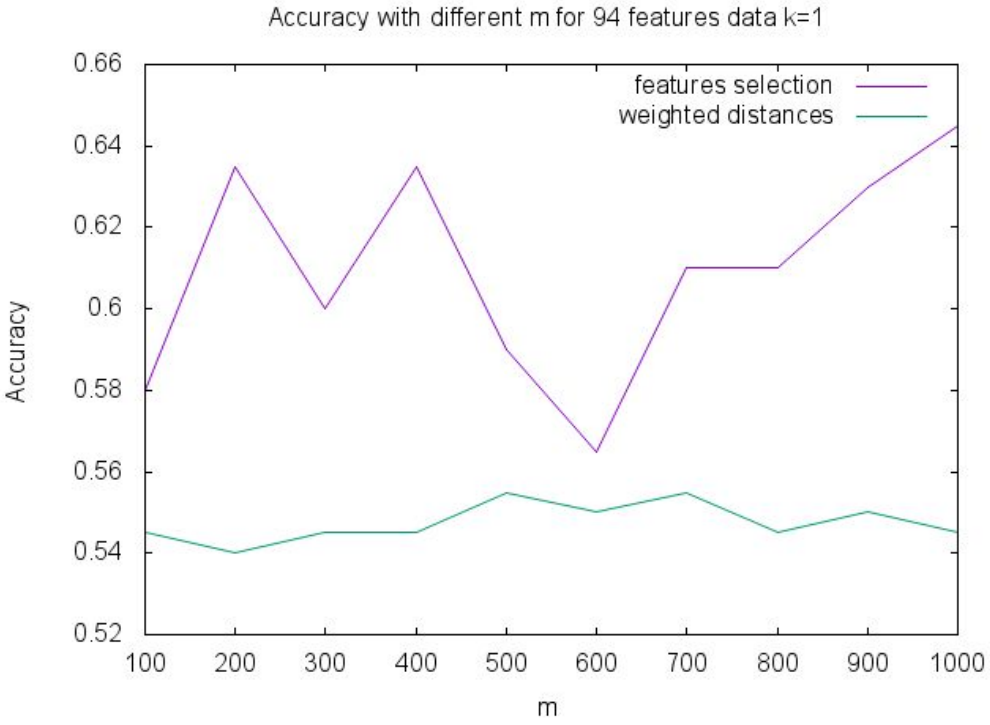
- **Analysis for 3.4.1**

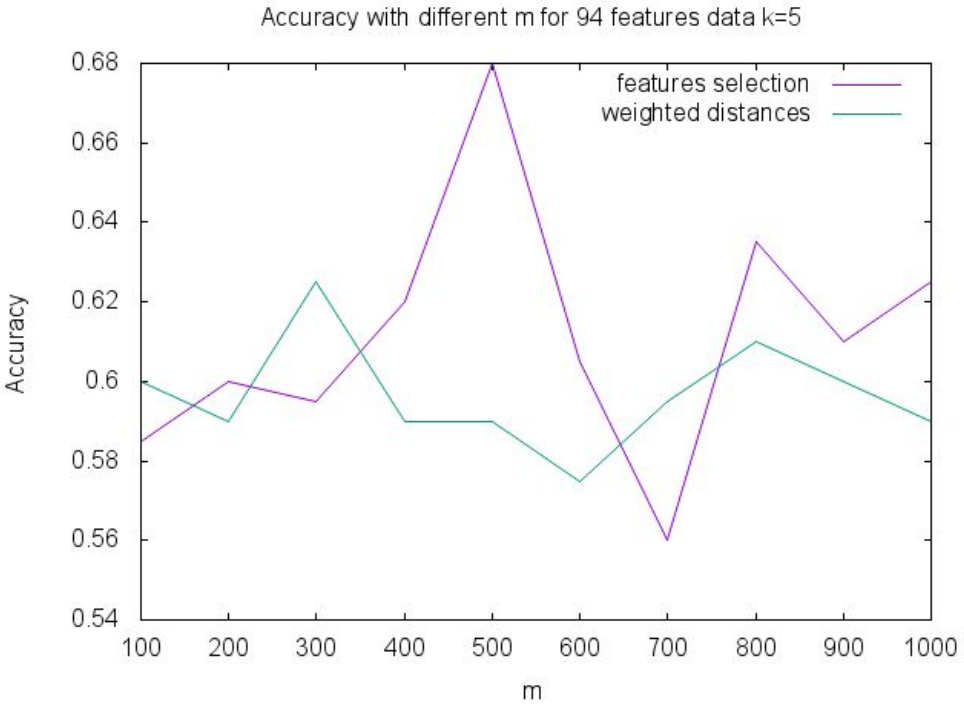




Based on the results of original KNN method, KNN with features selection relief and weighted distance relief improvements, we can conclude that KNN with relief functions do help us improve the accuracy of the accuracy. And for larger number of features, features selection relief improvement helps more for the accuracy. Weighted distance method basically has the same pattern with the original KNN method. The improvement of weighted distance method is not as such obvious as features selection method.

- **Analysis for 3.4.2**





According to the results for different m effect of 94 features dataset, we can conclude that weighted distance method is less sensitive to the change of m, and features selection method is

more sensitive to the change of m. For features selection method, it seems that larger m may not always lead to higher accuracy. For example when k = 5, m = 500 gives the best accuracy for the results, and m = 700 gives the worst accuracy. So when we do experiments, we need to choose a good m for better results.