Homework Assignment #4 – Part B

Use stratified 10-fold cross validation on the chess-KingRookVKingPawn.arff data set to compare naive Bayes and TAN. Be sure to use the same partitioning of the data set for both algorithms. Report the accuracy the models achieve for each fold and then use a paired t-test to determine the statistical significance of the difference in accuracy. Report both the value of the t-statistic and the resulting p value.

Accuracy achieved on each fold, from two models:

Fold Index	Naïve Bayes	TAN
1	0.871069182390	0.930817610063
2	0.864779874214	0.908805031447
3	0.896226415094	0.930817610063
4	0.877358490566	0.927672955975
5	0.886792452830	0.933962264151
6	0.861635220126	0.930817610063
7	0.902515723270	0.930817610063
8	0.889937106918	0.908805031447
9	0.883647798742	0.924528301887
10	0.877245508982	0.922155688623

Paired t-test result:

t-statistic = -9.5105
p-value = 5.425e-06

Alternative hypothesis: true difference in means is not equal to 0

The p-value is quite small, which means the difference in accuracies from these two models is significant.