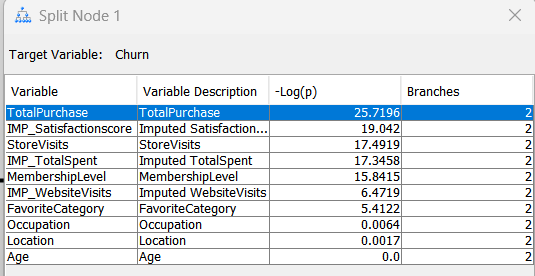
**Result and Analysis**

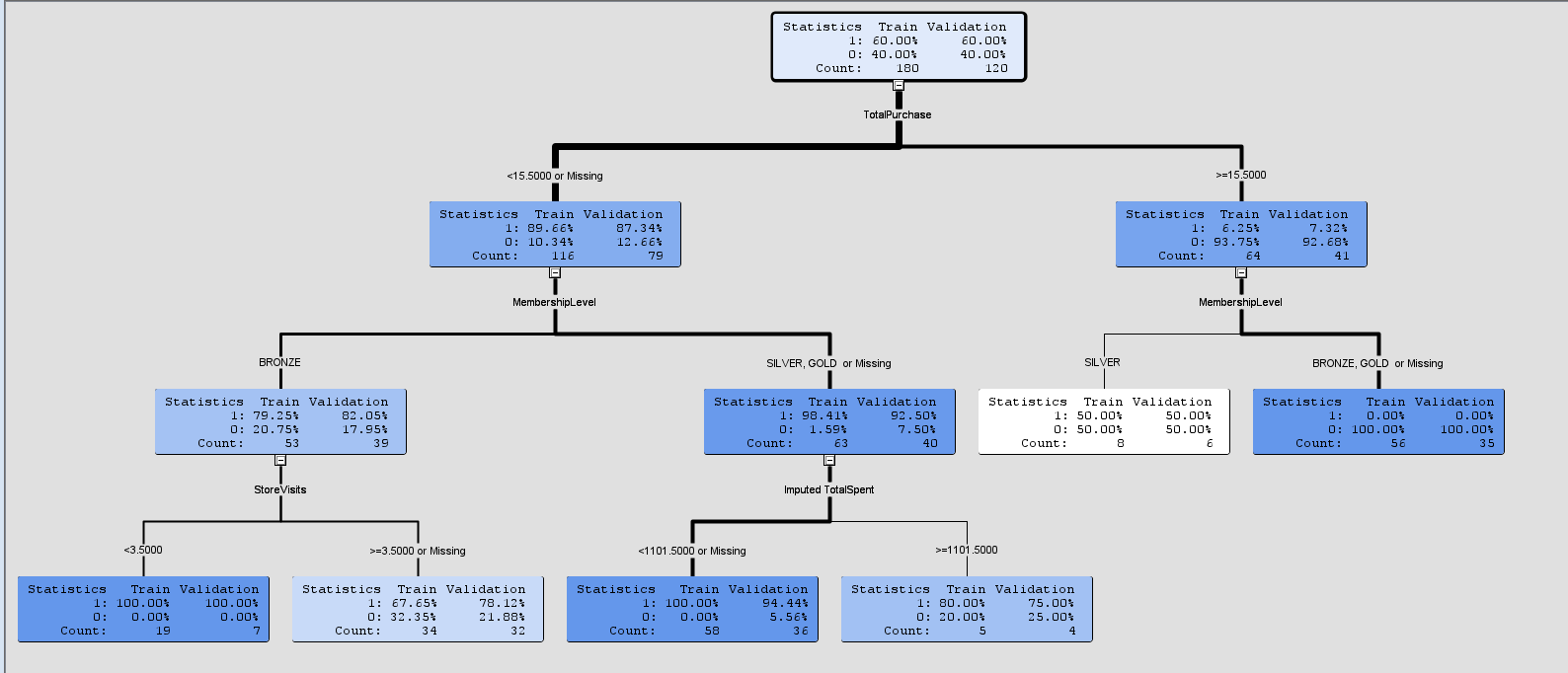
In summary, the results and analysis are retrieved from the main report.

The table below show the log worth of each variable that is being used to construct Decision Tree model. The higher the value means the better the variable to be used as a splitting node.



The total purchase is the best splitting variable to be used in the Decision Tree, so it should be

used as the parent node and followed by the other variables. This also provides insights that the number of purchases is related to the churn. The other important variables that affect churn are satisfaction score, total spent, store visits, favorite category, website visits and membership level. The age and location does not have much significance according to the table above.

Below is the maximal Decision Tree diagram which is the tree is grown until it is stopped by the splitting rule. The maximum depth set for the tree is 6.

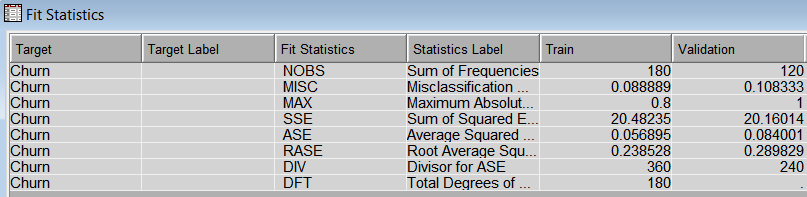
Below are the results. The total purchase with the most importance is placed as the parent node. When a client has total purchase more than 15.5, it will branch to the right and looking at the numbers, it is more likely that the client is not churned. Further split node with membership level, if the client is gold or bronze, it is very likely the client is not churned. However, if the client is silver, then it is about 50% that the client is churned. The node here might be an issue with the data size and we would expect silver and bronze to be classed together.

When the total purchase is less than 15.5 for a client, it is about 89% from the training data that the client is likely churned. This might mean that the client is not interested in the the products that they are browsing therefore less number of total purchase and eventually churned. When further splitting to membership level, it seems that all Bronze, Silver and Gold clients would have high probability to be churned.

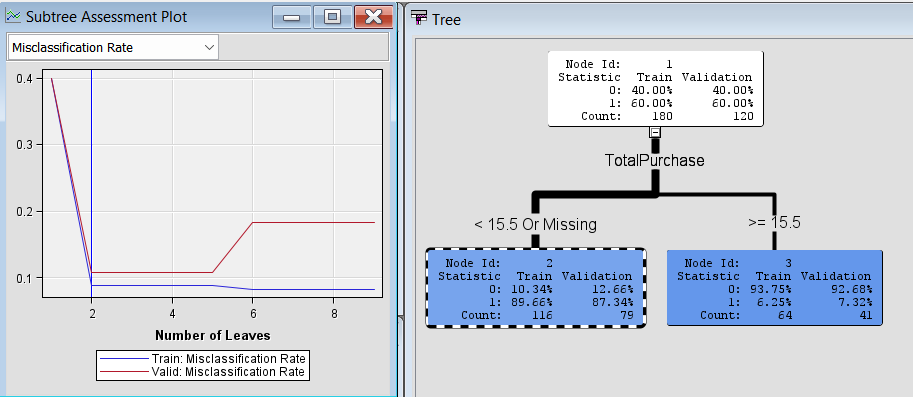
Under the Bronze node, it is further split into Store visits node where if the client is visiting less than 3.5 times, it is certain that the client is churned and if it is more than 3.5 times, then the client is less likely to be churned.

Under the Silver and Gold node, it is further split into Total spent where the spending is less than 1101.5, then it is a churned client. If the spending is more than 1101.5, then it is less likely for the client to be churned. Note that the decision tree below is a maximal decision tree and the decision tree is known to be prone to overfitting.

Below is the fit statistics of the Decision Tree above and it does not perform badly. If we focus on the mis classification rate on validation, we can observe that the rate is 0.108333 which is slightly higher than the rate 0.088889 from the train data.

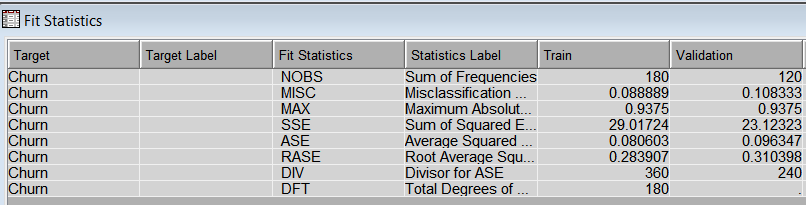


Another automatic pruning decision tree is ran and below is the results



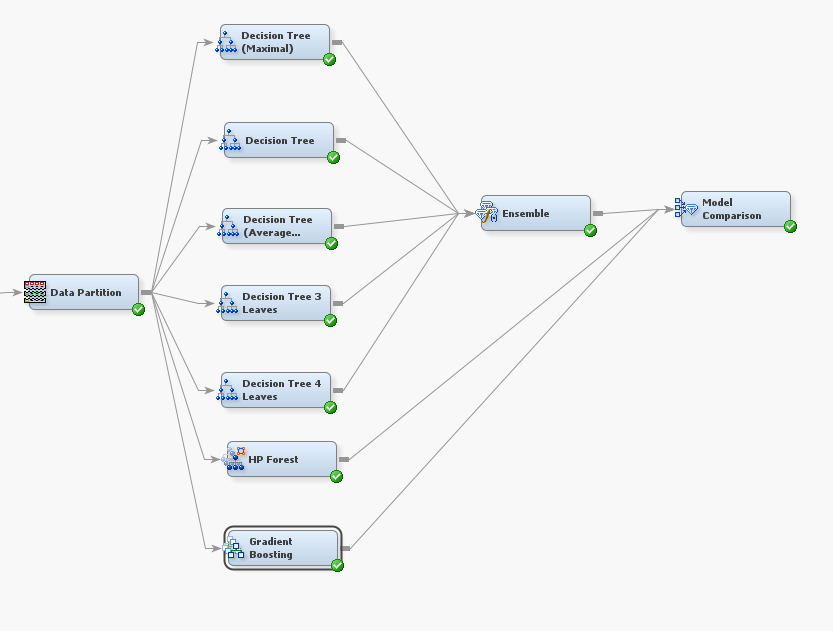
This decision tree diagram is a shallow decision tree with only 2 numbers of leaves. This can be explained that even though increasing the number of leaves does reduce the misclassification rate for training data, the misclassification rate increases in validation data when there are more than 5 numbers of leaves. SAS enterprise miner selects only 2 number of leaves with 1 branch to construct the model as it reaches the lowest misclassification rate for validation rate.

However, when compared the fit statistics of both models, the pruned model does not have better performance in terms of the error metrics.

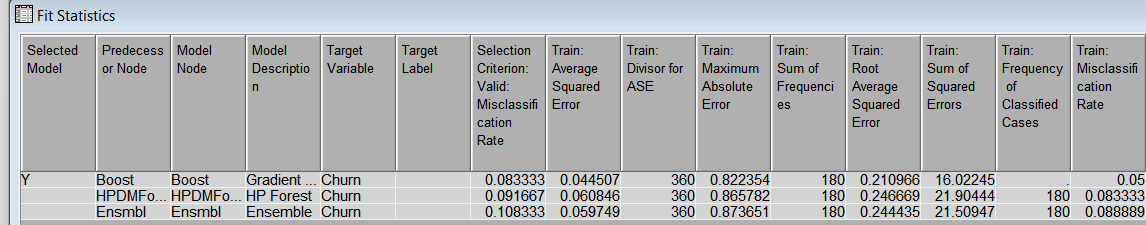


Therefore, bagging and boosting ensemble methods are required to further improve the models. For bagging, HP random forest and ensemble node is used while for boosting, Gradient Boosting node is used.

The overall diagram looks like below:



A model comparison is added to compare the performance of each models.



If we compare all the metrics here, the Gradient Boosting model is performing well compared to the other models. Especially on misclassification rate, it has the lowest of them all. This can be explained with the Boosting method is suitable for decision trees with shallow decision trees which is the case with decision tree models since our data set is small. Bagging method is usually applied and performed better when the dataset is large enough to ensure diverse subsets, also when the decision tree is deep which is helpful in reducing overfitting. However, that is not the case here as the decision trees constructed are shallow.

**Conclusion**

There are limitations with the data set that we have as the sample size is small and only variables like Total spent have high variable importance to the churn variable. Although other variables such as membership levels, satisfaction score, store visits and total spent variables have some variable importance, perhaps the sample size is not sufficient to capture the whole customer behavior. Gradient Boosting is seen to perform better than the single decision tree and the ensemble of decision trees as GB would improve the decision tree subsequently to get better results and proved to be efficient in shallow decision trees.

From the data analysis, we learned that clients that have lower total purchase and lower spending would have a very good chance to be churned. Therefore, we can filter out these clients and check the correlation with other variables such as the satisfaction score and their favorite category. This is to understand if they are feeling dissatisfied with our products and perhaps the category in our ecommerce site does not provide the expectation to their favorite category.

The store visits the client for records of visiting the physical store instead of shopping via e-commerce platform and this plays an impact variable worth to churn. We can perform deeper analysis to clients with a high number of store visits and their satisfaction score just to understand the correlation between the 2 variables. However, to have a more reliable model, more data sets are required in order to train the model. The age of the client and location of the client does not play much impact in affecting the churn variable.