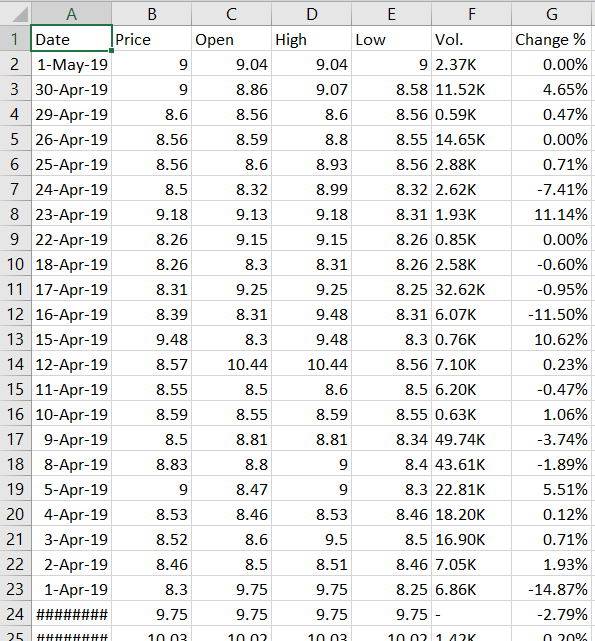
**MILESTONE 5: PREDICTIVE MODELING**

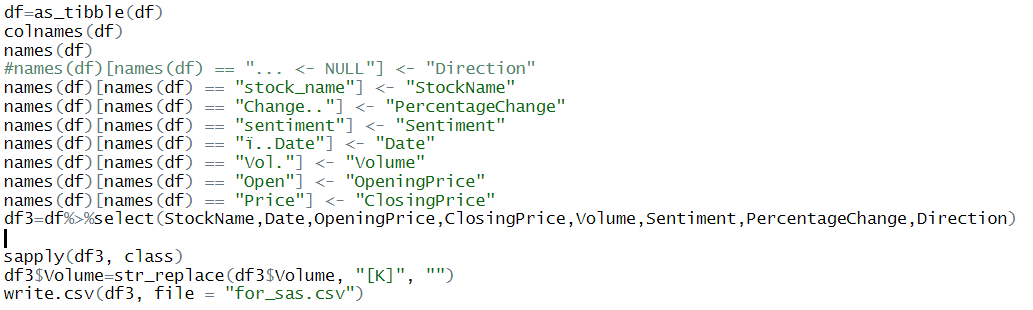
1.0 PRE-PROCESSING OF THE DATA

The stock data was crawled from investing.com for the stock name GEBHY or Genting Berhad. The stock data ranges from May 2014 to May 2019. The data contains column date, closing price, opening price, high, low, volume, and percentage change%. Preprocessing was conducted in R and feature engineering was conducted to get the direction of the stock based on the positive and negative value of the percentage change with ‘1’ being positive = direction up and ‘0’ being negative = direction down.

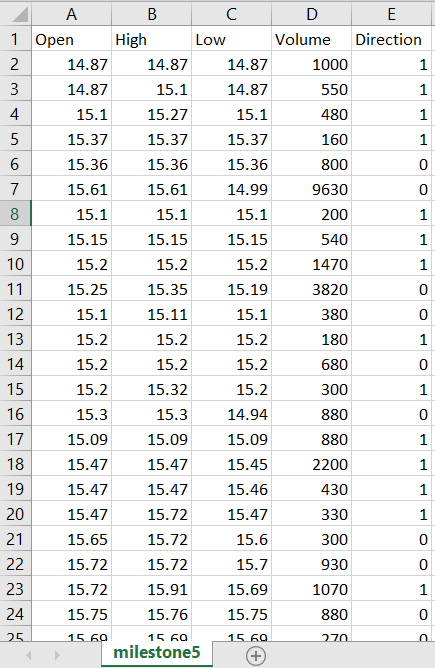


Preprocessing in R





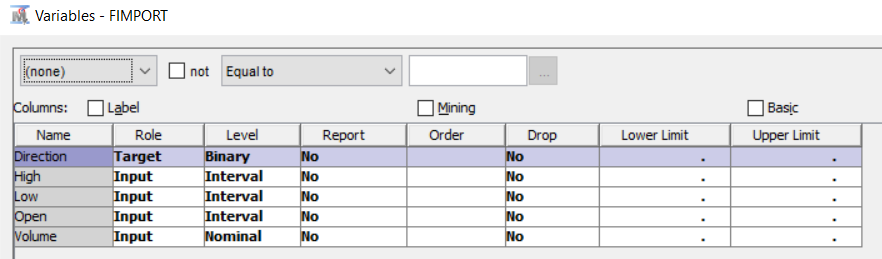
Processed data



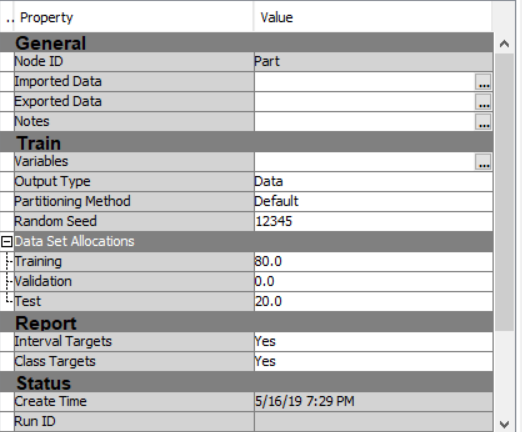
2.0 DATA LOADING AND MODELING IN SAS ENTERPRISE MINER

PROCEDURES:

1. Load data into SAS and set target variable



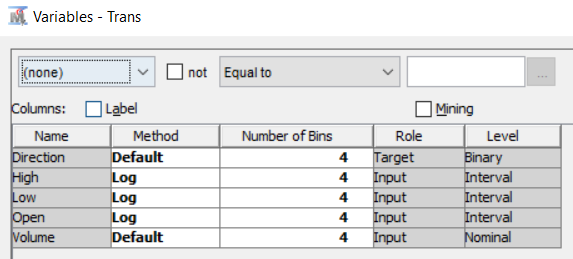
1. Data partitioned into 80:20 training to test set



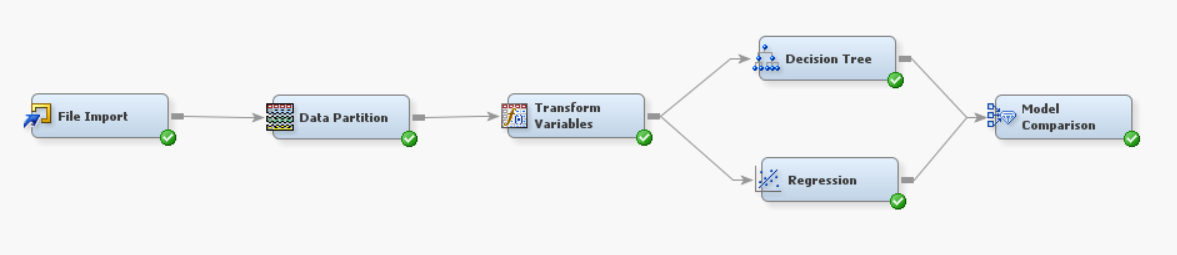
1. Interval attributes transformed

Skewed variables are transformed into a more modeling ready form

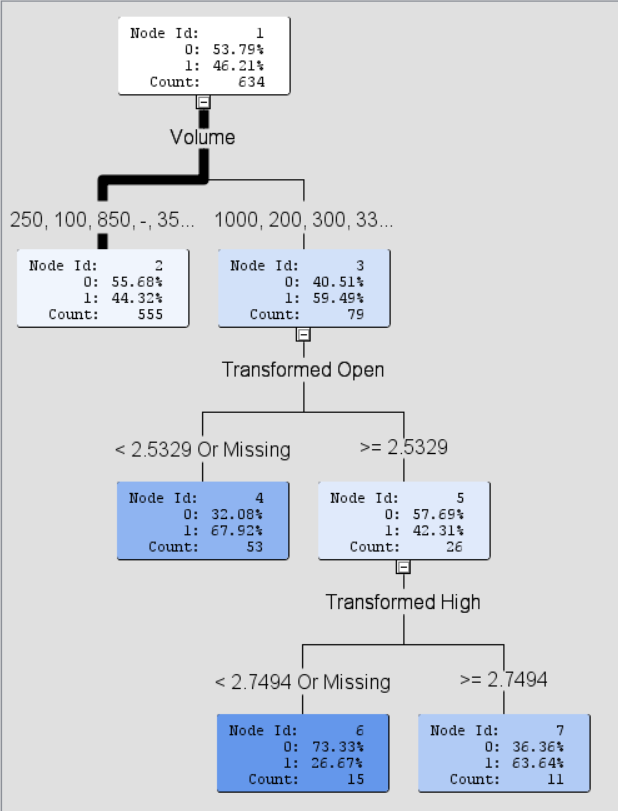




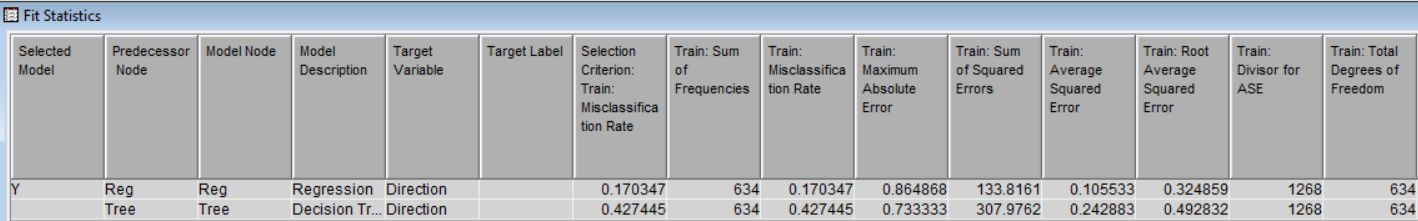
1. Decision tree and logistic regression models compared using the transformed data

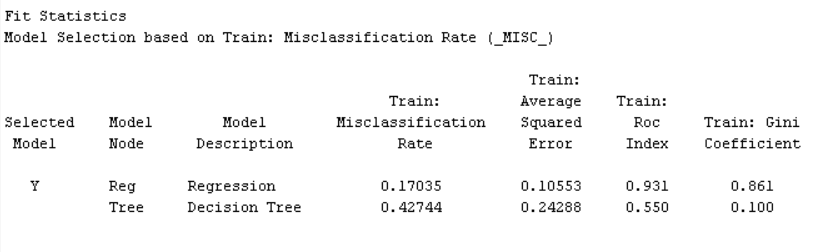


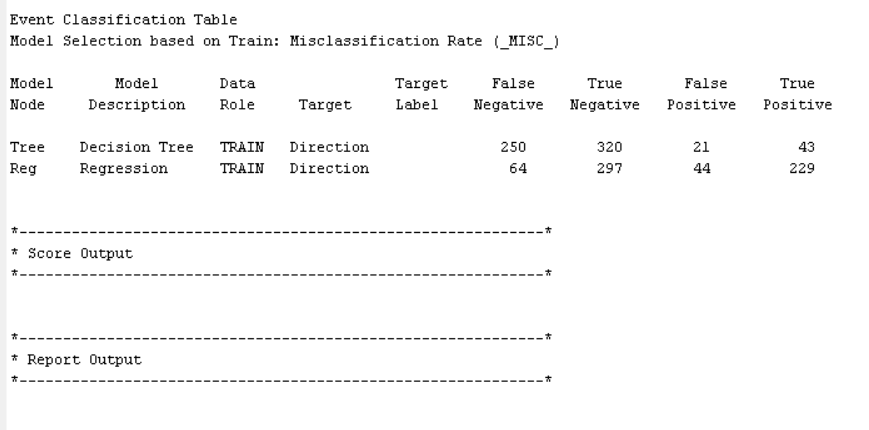
Decision tree



1. Model comparison to compare results between 2 models







The logistic regression model performed better than the decision tree model with lower misclassification of 0.17 against 0.42. The number of true positives were higher as well for the regression model with 229 TP vs 43 TP of the decision tree model. Insights from this milestone is that the regression model is better overall to predict the direction of the stock based on the crawled data we have compared to the decision tree model and it can be a guide for stock monitoring to know what the direction will be with an acceptable accuracy rate.