

## Results of RandomForest Modeling

Last amended: 28<sup>th</sup> Oct, 2019

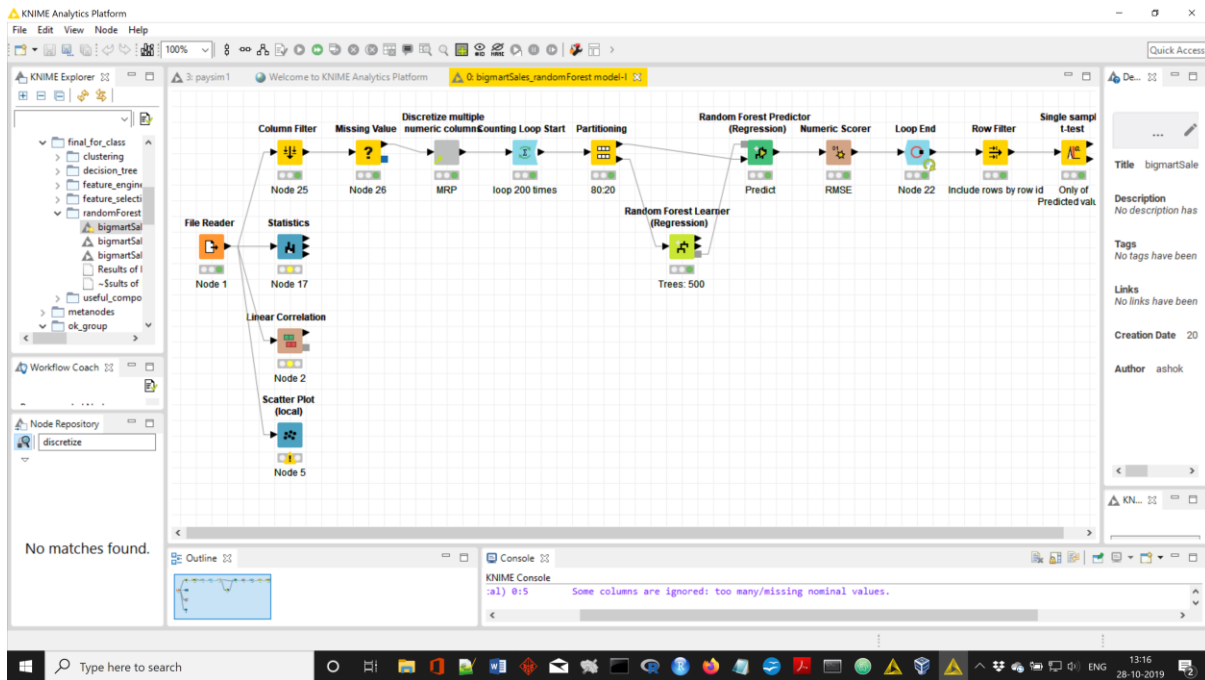
My folders: D:\data\OneDrive\Documents\knime-workspace\final\_for\_class\randomForest

D:\data\OneDrive\Documents\big\_mart\_sales\_problem

<https://datahack.analyticsvidhya.com/contest/practice-problem-big-mart-sales-iii/>

### 1. bigmartSales\_randomForest model-I

- a. Fill in missing values
- b. Discretize MRP



**Single Sample T-Test**

**Descriptive Statistics**

	N	Missing Count	Mean	Standard Deviation	Standard Error Mean
Prediction (Item_Outlet_sales)	200	0	1,164.1673	8.752	0.6189

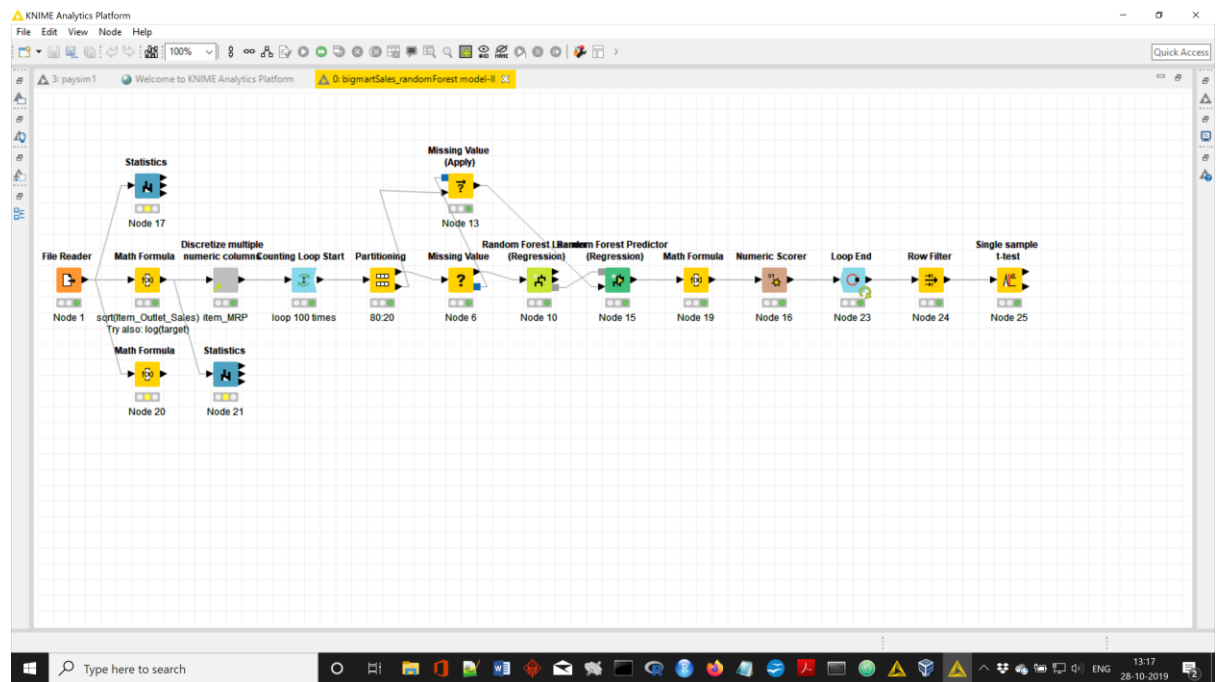
**Single Sample Test**

Confidence Interval (CI) Probability: 95.0%

	Test Value	t	df	p-value (2-tailed)	Mean Difference	CI (Lower Bound)	CI (Upper Bound)
Prediction (Item_Outlet_sales)	0.0	1,881.1495	199	0.0	1,164.1673	1,162.9469	1,165.3876

Figure 1: Note that this result gives the least std error of mean even though Mean is highest.

2. bigmartSales\_randomForest model-II
  - a. Discretize MRP
  - b. Missing values filled in for 80% and use the model for rest 20%
  - c. Normalize outputsales using sqrt()



missing value (Apply)

Test statistics - 5:25 - Single sample t-test

File

Single Sample T-Test

Descriptive Statistics

	N	Missing Count	Mean	Standard Deviation	Standard Error Mean
prediction(Item_Outlet_Sales)	200	0	1,131.6415	28.1243	1.9887

Single Sample Test  
Confidence Interval (CI) Probability: 95.0%

	Test Value	t	df	p-value (2-tailed)	Mean Difference	CI (Lower Bound)	CI (Upper Bound)
prediction(Item_Outlet_Sales)	0.0	569.0393	199	1.54E-321	1,131.6415	1,127.7199	1,135.5631

Figure 2: An improvement over result of 1st model. Note the limits of confidence interval. Its limits are less than Model 1. It is the best model.

3. bigmartSales\_randomForest model-III
  - a. Generate missing values using predictive analytics
  - b. Also normalize outputsales and discretize MRP

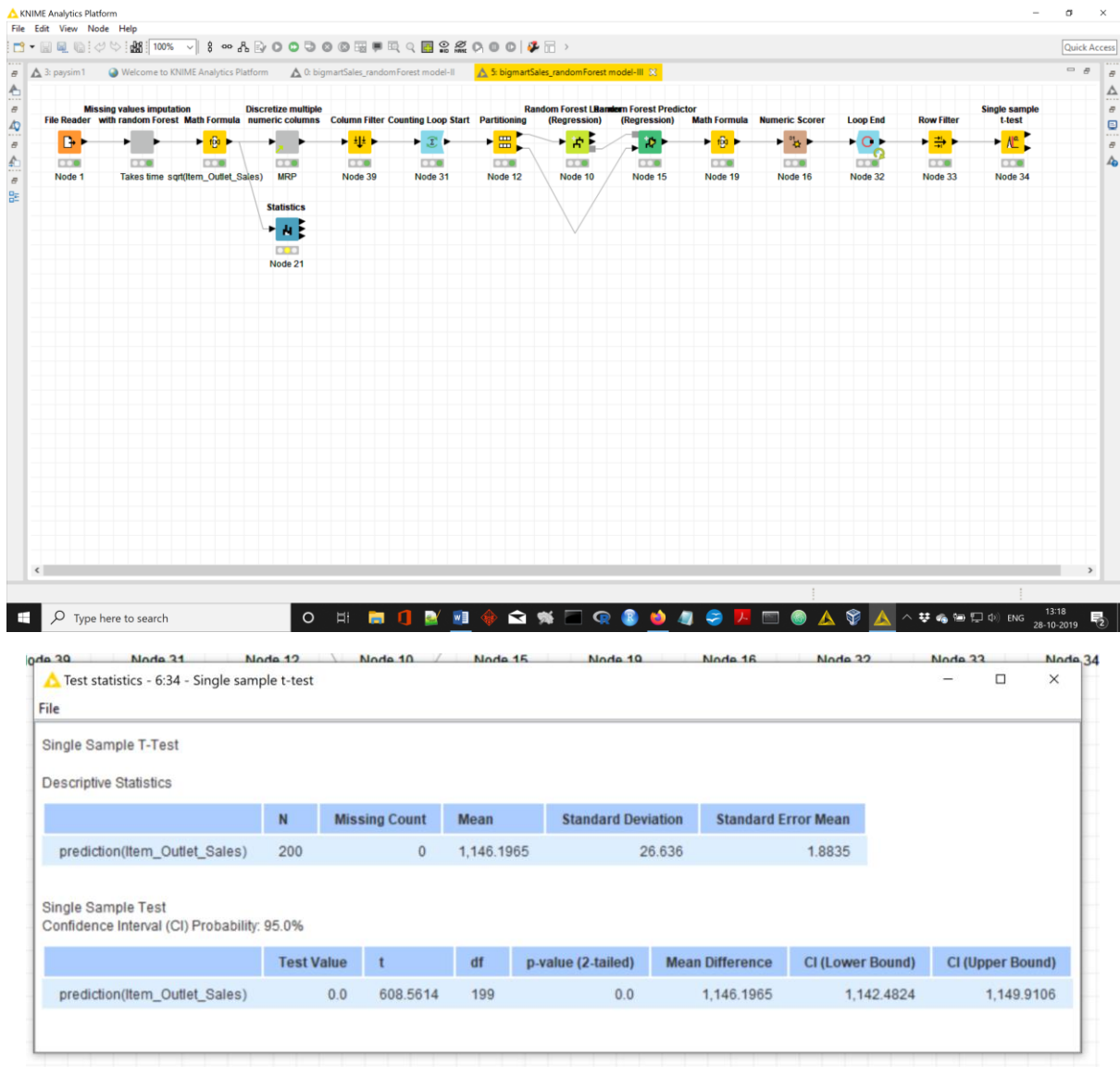


Figure 3: This model falls in between Model 1 and model 2

\*\*\*\*\*