SAS® GLOBALFORUM 2015

The Journey Is Yours

Gross Margin Percent Prediction: Using the Power of SAS® Enterprise Miner™ 12.3 to Predict the Gross Margin Percent for a Steel Manufacturing Company

Kushal Kathed[1] , Ayush Priyadarshi[2] , Dr. Patti Jordan[3]
[1,2]Graduate Student, Oklahoma State University
[3]Visiting Assistant Professor, Oklahoma State University

SAS®GLOBALFORUM

Gross Margin Percent Prediction: Using the Power of SAS® Enterprise Miner™ 12.3 to Predict the Gross Margin Percent for a Steel Manufacturing Company

Kushal Kathed[1], Ayush Priyadarshi[2], Dr. Patti Jordan, Visiting Assistant Professor at Oklahoma State University
[1,2]SAS & OSU Data Mining Program, Oklahoma State University

Abstract

Predicting the profitability of future sales orders in a price sensitive highly competitive make-to-order market can create a competitive advantage for an organization. Order size and specifications vary from order to order, customer to customer and may or may not be repeated. While it is the intent of the sales groups to take order for a profit, because of the volatility of steel prices and the competitive nature of the markets, gross margins can range dramatically from one order to the next and in some cases be negative.

Understanding the key factors affecting the gross margin percent and their impact will help the organization can reduce the risk of non-profitable orders and at the same time improve their decision making ability on market planning and forecasting. The objective of this paper is to identify the best model amongst multiple predictive models inside SAS® Enterprise Miner, which could accurately predict the gross margin percent future orders.

Data description

The data used for the project consisted of over 30,000 transactional records and 33 input variables. The sales records have been collected from multiple manufacturing plants of a leading steel manufacturing company. Variables such as order quantity, Customer location, Sales group, as well as others were used to build predictive models. The target variable gross margin percent, is the net profit on the sales considering all the factors such as labor cost, cost of raw materials, etc.

Data Preparation

Dummy variables for all categorical variables such as area, country, sales office, plant, etc. were created using SAS® Enterprise Miner 12.3. The data was partitioned using the data partition node in 60:40 ratio for training and validation respectively to carry out further analysis . Next, DMDB and Stat Explore node were used to understand the distribution of the variables. The average gross profit percent for the steel manufacturing company is 17.21.

Exploratory Analysis

Exploratory analysis on the data reveals that few sales groups such as 8 and 34 showed extraordinary performance while other groups such as 57 and 48 performance was had least mean gross margin percent.

Countries like Costa Rica and Israel have negative average Gross Margin Percent. This means most of the deals over there were in loss.

Below map build using GMAP procedure shows the distribution of Gross margin percent distribution across the United States. South Carolina had the least gross margin percent and whereas North Dakota had highest Gross Margin Percentage.
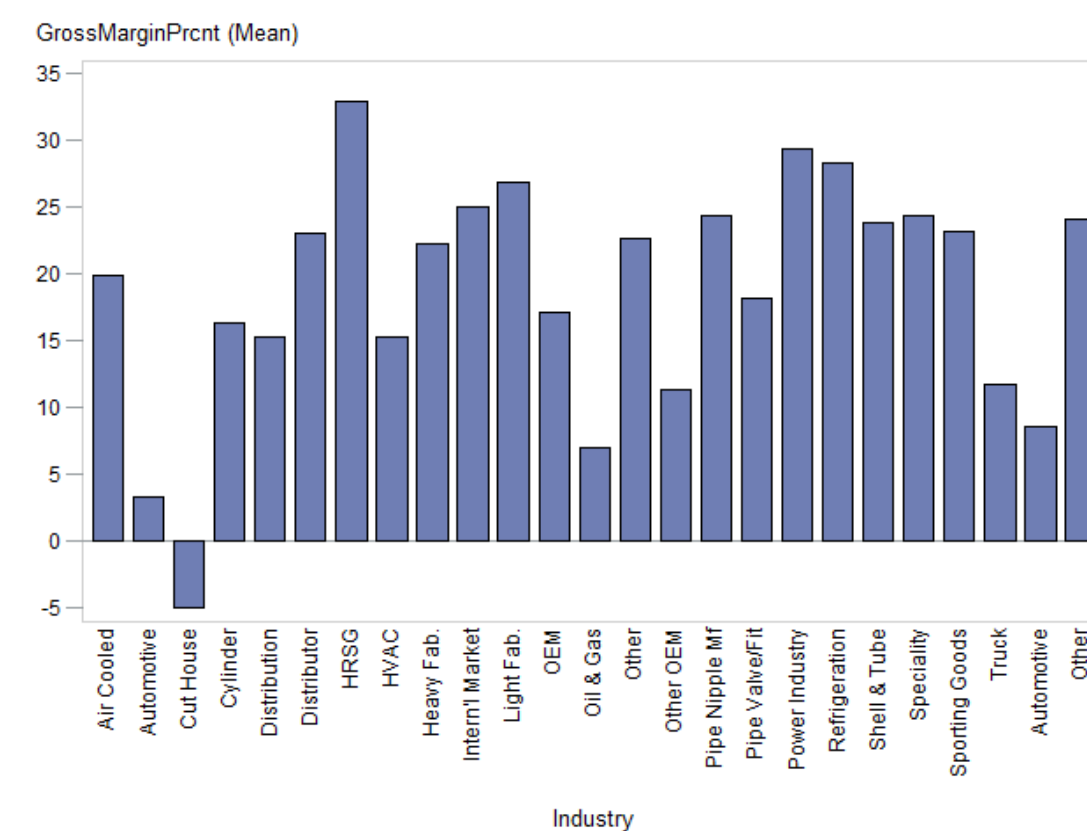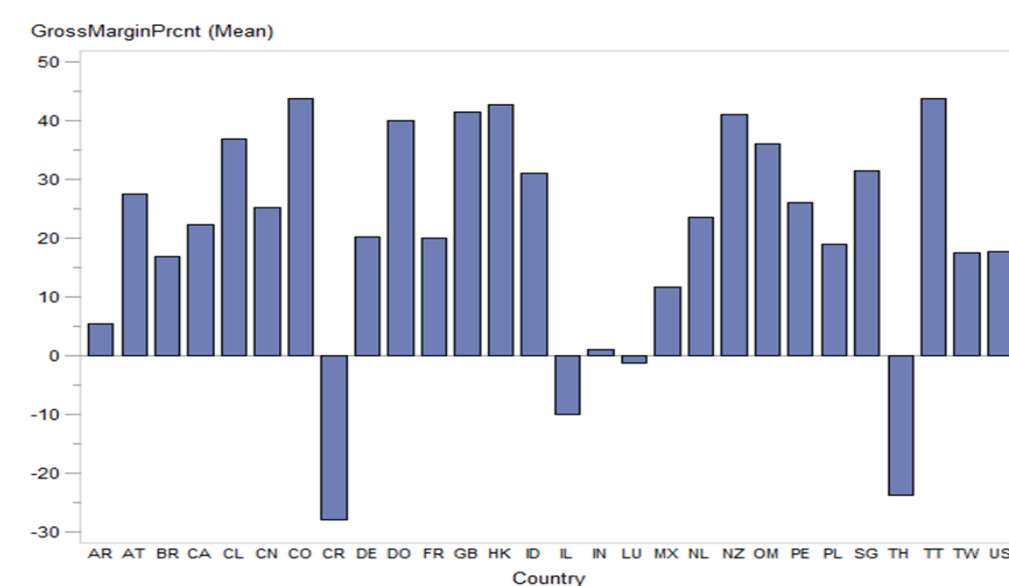

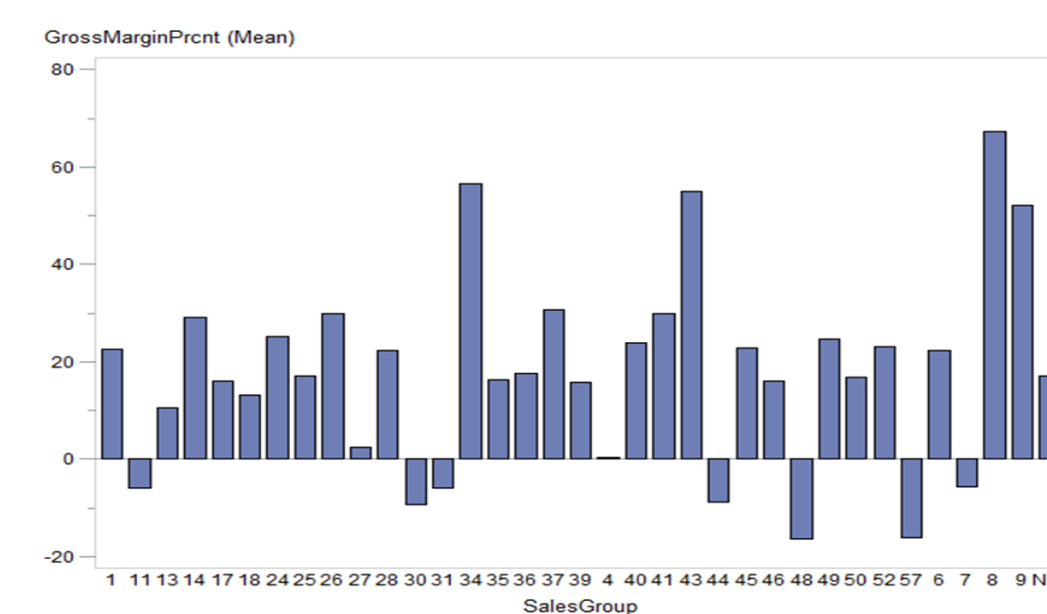Fig. 1. Average Gross Margin vs. Industry

Fig. 2. Average Gross Margin vs. Sales Group

Fig. 3. Average Gross Margin vs. Country

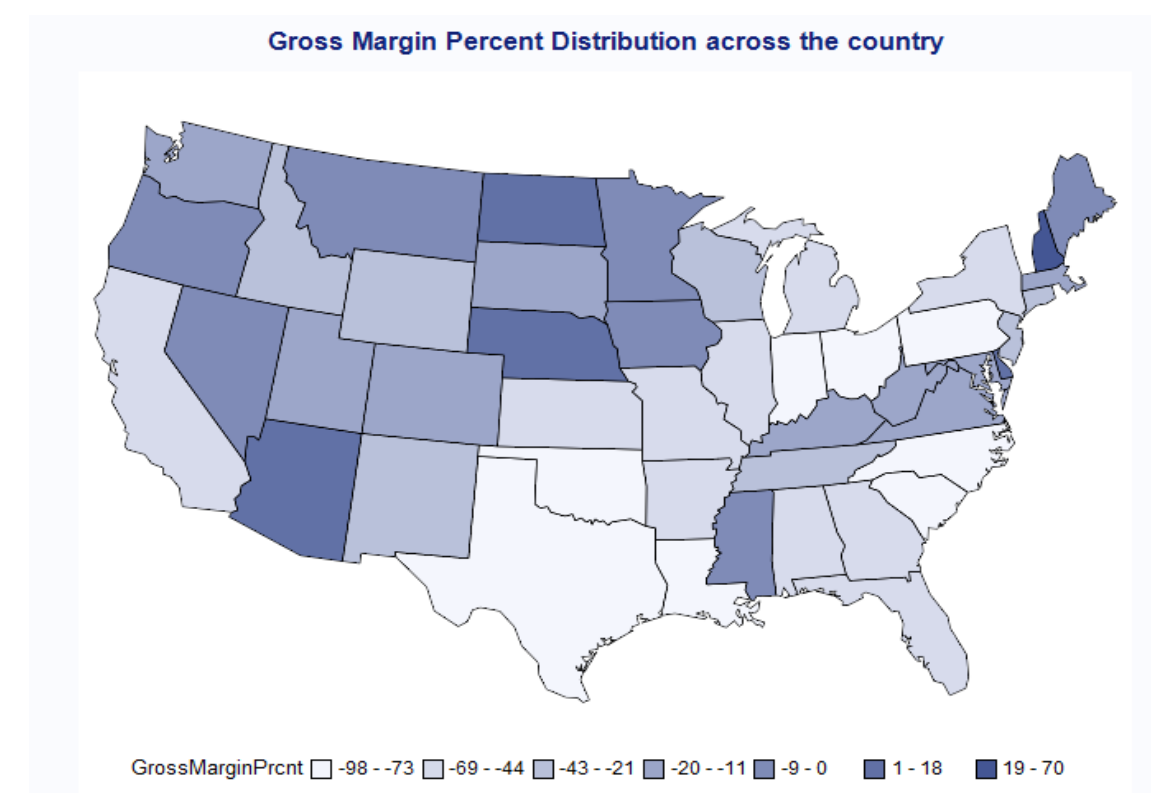Fig. 4.Average Gross Margin vs. Area

Predictive Modeling

Data mining models such as Decision Tree, Neural Network and various variations of Regression were built using Enterprise Miner 12.3. Finally, an ensemble model which is a combination of all the predictive models was built to predict the target. These models were later compared using the Model Comparison node in order to evaluate the best performing model using validation Average Square Error as the selection criteria.
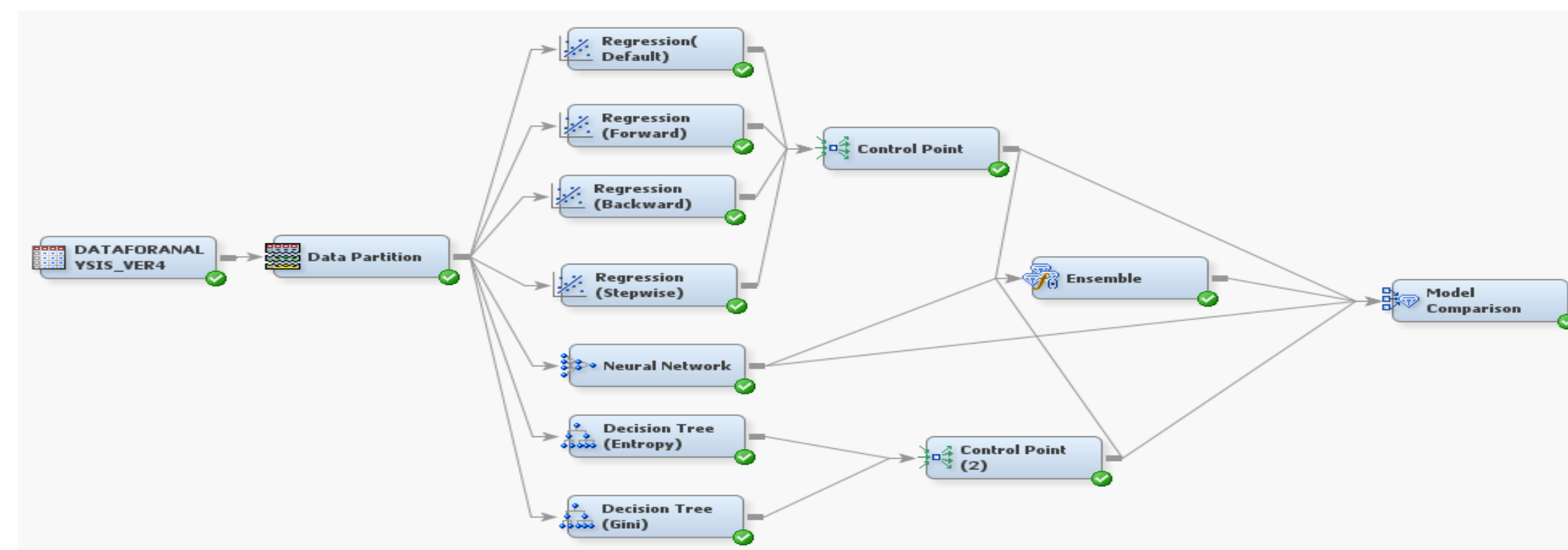

Fig. 5. Model Diagram (Nodes)

Model Comparison

Selected Model	Model Node	Model Description	Selection Criterion: Valid: Average Squared Error	Train: Average Squared Error	Train: Target Variable
Y	Ensmbl	Ensemble	12.55435	10.42199	GrossMarginPrcnt
	Reg3	Regression (Backward)	18.45173	16.06888	GrossMarginPrcnt
	Reg	Regression(Default)	18.454	16.06326	GrossMarginPrcnt
	Reg2	Regression (Forward)	18.4693	16.0789	GrossMarginPrcnt
	Reg4	Regression (Stepwise)	18.4693	16.0789	GrossMarginPrcnt
	Tree	Decision Tree (Entropy)	18.76026	14.62833	GrossMarginPrcnt
	Tree2	Decision Tree (Gini)	18.76026	14.62833	GrossMarginPrcnt
	Neural	Neural Network	21.94406	16.94338	GrossMarginPrcnt

Table 1. Model comparison table

On the basis of validation average square error, Ensemble model comes out to be the best out of all the predictive models. The validation average square error of Ensemble model is 12.55 whereas for the training dataset it is 10.42.

Results

- Stat Explore node provides us the worth estimate of the variables in order to predict the target variable. Variables such as Sales Group, Usage, Sub Division, Plant, Product, Industry, Footage, Profit Center and Raw Materials are the important variables that affect gross margin percentage.
- Among the significant plants, plant 10, 11, 30 and 31 have positive correlation while the plant 20, 21, 22, 41, and 43 have negative correlation with the target variable (gross margin percent).
- Among the significant industries, HVAC and Heavy Fab. are positively correlated while industries such as air cooled, automotive, cylinder have negative correlation with the target variable.

References

- http://www2.sas.com/proceedings/sugi29/251-29.pdf
- http://support.sas.com/resources/papers/proceedings14/1784-2014.pdf

Acknowledgements