Profitability Predictions Report

Multiple Regression in R

Ruthesan Balasubramaniam | Data Analyst / Big Data | 19/10/2018

Introduction:

The problem in this task is to predict the sales volume of new electronics based on the existing product details and its reviews. This data set contains new values such as Competitors, Professional and Age, these are completely new attributes comparing with our previous task in rapid miner. Also assessing the impact services reviews and customer reviews have on sales of different product types.

- − This task contains to data sets named Existing products having data of products we sold and New products are the products we are planning to add in our inventory for future sales.

- − The existing products contains 245 sample values and 21 attributes which are product type, ID, price, rating, dimensions, service reviews, selling rank, profit, volume and type of customers who bought these products. This data set having 16 missing values mostly in bestselling rank attribute.

- − The new products data set have 24 samplings and same attributes from the above data set where the volume column has to be predicted.
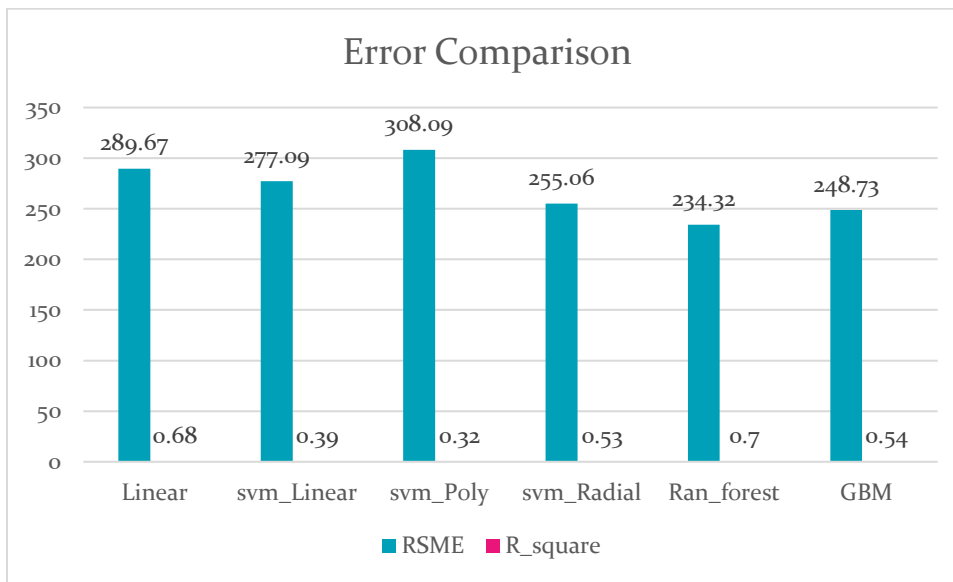
Preprocessing:

- • The existing products data set was analyzed, and the data types were converted into numeric values other than the product type. This factor data of product type is converted into dummies. And the best seller ranking row was removed. The below shown correlation matrix was drawn.
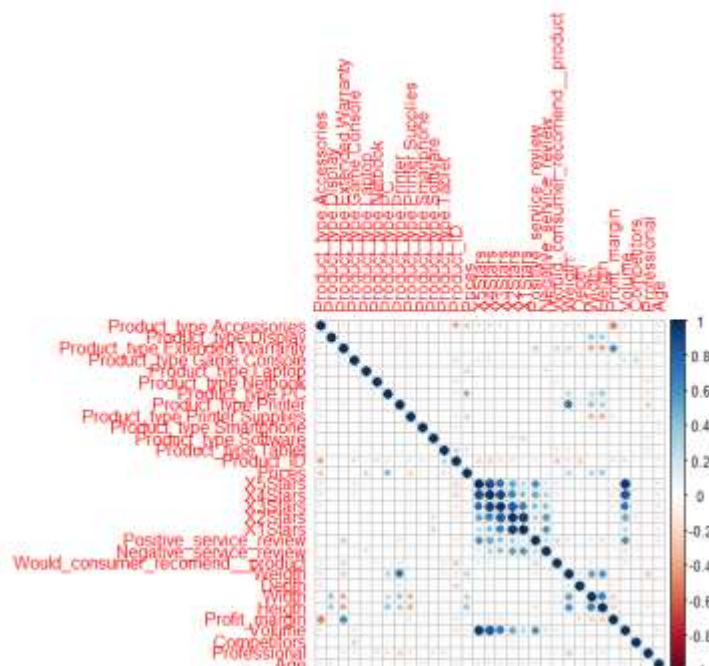
- Outliers of each attributes row removed from the data set.
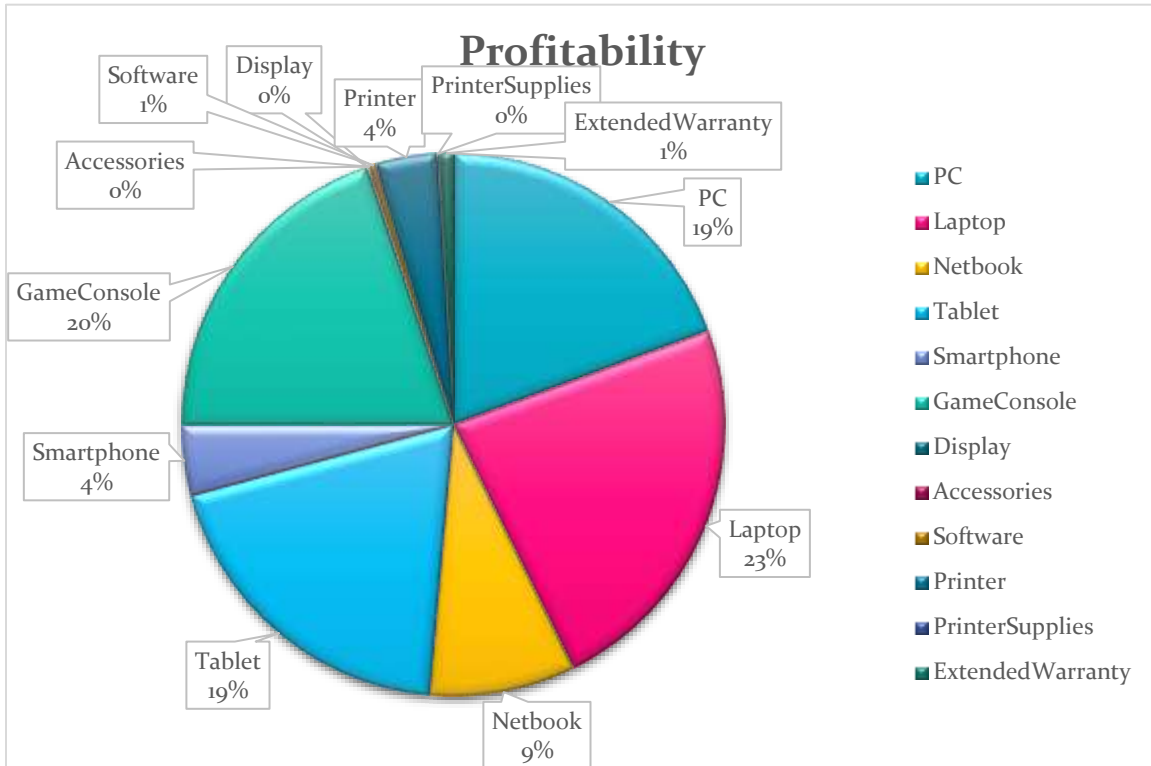- Attributes have high correlation with product volume are selected and its colinear attributes removed.

Modelling:

- The preprocessed data sliced into training and testing set with the ratio of 75:25.

Error Comparison

Model	RSME	R_square
Linear	289.67	0.68
svm_Linear	277.09	0.39
svm_Poly	308.09	0.32
svm_Radial	255.06	0.53
Ran_forest	234.32	0.7
GBM	248.73	0.54

- From the above plot came to know that the Random Forest model producing less error comparing with other methods, so this model is taken for the volume prediction for the new products data set.

Sales Prediction:

Profitability

Software 1%	
Display 0%	
Accessories 0%	
Printer 4%	
PrinterSupplies 0%	
ExtendedWarranty 1%	
PC 19%	
GameConsole 20%	
Smartphone 4%	
Laptop 23%	
Tablet 19%	
Netbook 9%	

Legend:
- PC
- Laptop
- Netbook
- Tablet
- Smartphone
- GameConsole
- Display
- Accessories
- Software
- Printer
- PrinterSupplies
- ExtendedWarranty

- The top 4 profitable product types are Laptops, Game consoles, PCs and Tables Respectively.