

Data mining approach for Sales Prediction

Rathnadiwakara, A.S.K. and Liyanage, S.R.

Department of Statistics and Computer Science, University of Kelaniya

Department of Software Engineering, Faculty of Computing and Technology, University of Kelaniya

Email: asrathnadiwakara@gmail.com, sidath@kln.ac.lk

Abstract

Nowadays predictive analysis is more popular among companies to improve their business profits. Those companies differentiate what they do from “Data Mining”. The characteristic deduction is that data mining is limited to the discovery of patterns, whereas predictive analytics allows the application of the patterns to new data to predict unknown values. The main aim of data mining is to extract knowledge from the data at hand, increasing its intrinsic value and making the data useful.

Today most business areas using many strategies to improve their business profits. They are mostly use traditional methods. Therefore, the company’s efficiency and profitability, goes to the critical situation. When considering today’s business arena, the most important things are good efficiency and correct strategies for a business. Converting to the new technologies companies can achieve their business goals and they can reveal their sales life-cycle.

This research proposes for a medium scale tyre dealing company which is situated in Colombo. It is important to company to accurately predict their future order details and salary income before having unprofitable occasion. Company could conduct sufficient stock when talking prediction support. That is the best solution to reduce time, save importing cost, growth for income and manage resources.

Data mining algorithms and techniques used for the prediction process and used MS SQL Server 2008 R2 with Analysis Server and Business Intelligent Development Studio for modeling process. Analysis Services contains number of standard data mining algorithms. Decision Tree, Neural Network and Clustering data mining models were attempted for the prediction. Decision Tree is a graph of decisions and their possible consequences, represented in form of branches and nodes. A Neural Network is a parallel distributed processor that has a propensity for experiential knowledge and making it available for users. Clustering is used to place data elements into related groups without advance knowledge of the group definitions. The best algorithm was selected for each model and it focused on five main attributes which were referred to as factors affecting a sales process such as Item Code, Item Type, Item Quantity, Item Value, Item Sold Date, etc. variables were used in data mining process. Among those variables five variables were selected for the mining process.

Dataset arranged with 30% data for testing process and 70% data for the training process. According to the predicting probabilities, Decision Tree algorithm were performed 99.53%, Neural Network

algorithm were performed 73.36% and Clustering algorithm were performed 67.79%. Clustering model belongs to the lowest predict probability value. Therefore Clustering model was the worst model. Decision Trees model contains highest predicted value 99.53%. Therefore it was the best model. Neural Network model was also a good model. The Score results indicate that Decision Trees mining model has the best score 1.00 and followed by Neural Network mining algorithm with score of 0.92 and clustering mining algorithm with 0.94. Considering the data mining lift chart for mining structures, it graphically represents the improvement that a mining model provides when compared against a random guess, and measures the change in terms of a lift score. By comparing the lift scores for various portions of the dataset and for different models. According to the Lift chart representation, Decision Trees curve present in upper in the chart with comparing other carvers. By considering lift chart, score and target population with predicting probabilities, Decision Trees algorithm was the best one for prediction process.

Finally, Data mining model was implemented using Decision Trees algorithm. According to these predicting results, the company can handle their imports optimizing the available resources; storage, time, money. Therefore this research would benefit the Company to improve their incomes.

Keywords: *Data Mining, Decision Tree, Neural Network, Clustering*