

Data Analysis in Business Intelligence for Mathematical Prediction of Online Sale

Varsha B. Saknure¹, D. V. Jadhav²

¹Zeal College of Engineering and Research, Narhe, Pune 411041

²Assistant Professor, Zeal College of Engineering and Research, Narhe, Pune 411041

Abstract: *Business-intelligence is an effective means of improving the decision making within an organization applied to profit making private companies. The online purchase and trading data available on website can be a key reference to analyze for the decision making in small scale business. The electronics product market is more volatile in nature and is having frequent variation. Such behavior of product variation has increases the risk of over inventory towards the dealers. The effective analysis of behavior of the product sale over the year can be set as a foundation for decision making. The recent BI tools are used to analyze the data to develop the report generation of the sale. These reports are studied to develop the thumb rules with respect to the variation of sale. Such thumb rule is designed to use as a knowledge base in developing a mathematical model for the predictions of sales behavior of electronics products. The Prestashop based sales module is used along with the BI tool Pentaho for the development of sales report in this paper. The MATLAB is used to develop the mathematical prediction.*

Keywords: Business Intelligence, Sale of Electronics products, Sales Module, Pentaho, Mathematical Prediction

1. Introduction

Business intelligence (BI) continues to be a top priority for many organizations, and the promises of BI are rapidly attracting many more proponents. Organizations are grappling to make sense of the rapidly increasing volume, velocity, and variety of data generated by both internal and external resources. Consequently, BI has become a critical foundation of competition for many organizations and has consistently been ranked among the top two agenda items of senior executives over the last several years. Business intelligence continuously ranks as one of the most widely searched terms on gartner.com and remains a topic of interest in academic and practitioner research. One overarching theme that has surfaced in the research is that the BI used within an organization must suit the problem space, or decision environment, within which it is used and that this match is key to BI success. However, practitioner-oriented publications and academic research suggest that this success has not yet been realized in many organizations and that the users do not necessarily make the connection between their BI capabilities and the decision environment.

BI capabilities are critical functions that help an organization improve both its adaptation to change its performance. Although BI capabilities have been studied from organizational and technological perspectives, some organizations fail to achieve BI success. This may be because the relationship between the decision environment and BI capabilities has remained largely unexamined. Examining this relationship is important, however, because the primary purpose of BI is to support decision making in organizations. The purpose of this project is to provide a better understanding of BI success by examining the impact of BI capabilities on BI success in the presence of different decision environments

2. Literature Review

Wingyan Chung, Tzu-Liang (Bill), Tseng [1] has studied Discovering business intelligence from online product reviews: A rule-induction framework, as per his paper the Online product reviews are a major source of business intelligence (BI) that helps managers and marketers understand customers' concerns and interests. The large volume of review data makes it difficult to manually analyze customers' concerns. Automated tools have emerged to facilitate this analysis, however most lack the capability of extracting the relationships between the reviews' rich expressions and the customer ratings. Managers and marketers often resort to manually read through voluminous reviews to find the relationships. To address these challenges, we propose the development of a new class of BI systems based on rough set theory, inductive rule learning, and information retrieval methods. We developed a new framework for designing BI systems that extract the relationship between the customer ratings and their reviews. Using reviews of different products from Amazon.com, we conducted both qualitative and quantitative experiments to evaluate the performance of a BI system developed based on the framework.

Jie Gao et.al [2] has studied the online purchase decision making: The effects of unconscious thought, information quality, and information quantity. As per this literature, the prosperity of online shopping has led e-commerce vendors to provide increasingly rich information, particularly for experience products, to enhance consumers' shopping experience and satisfaction. However, there is little awareness that consumers may not be able to process all the information available because of human beings' limited information processing capacity. Online shoppers could be easily confused when facing rich information, particularly when the amount of information greatly exceeds their processing capacity.

Sahar Karimi, K et.al [3] has studied The effect of prior knowledge and decision-making style on the online purchase decision-making process: A typology of consumer shopping behavior. This paper provides an empirical typology of online decision-making purchasing behaviour. The study explores how the online purchase process is affected by individual decision-making style and knowledge of product. Drawing from the decision analysis and consumer behaviour literatures, we present a typology of online purchase decision-making behaviour and introduce four archetypes of online consumers. A number of experiments have been conducted in two online settings: retail banking and mobile networks. Based on an extensive video analysis, we have captured four process-related dimensions (number of cycles, duration, number of alternatives and number of criteria) using a business process modelling approach. Significant differences in all process-related dimensions were found across the four archetypes. The study improves understanding of the different types of online consumers and their process outcomes. The findings are useful for online retailers seeking to improve the way they support the four archetypes of online shoppers throughout the decision-making purchasing process.

R. Eric Hostler et.al [4] has studied Assessing the impact of recommender agents on on-line consumer unplanned purchase behavior. The Recommendation agents (RAs) have been used by many Internet businesses such as Amazon and Netflix. However, few authors have studied how consumer behavior is affected by those that make suggestions to online consumers based on their recent shopping behavior. Fewer still have examined the role that RAs play in influencing impulse purchasing decisions online. Our study developed a theoretical model to illustrate the impact of RAs on online consumer behavior. The model was tested through an online shopping simulation which used a collaborative filtering based product RA. Particular attention was paid to the effects of an RA on consumer behavior; we found that it increased promotion and product search effectiveness, user satisfaction with the website, and unplanned purchases. Results showed that the model provided insights into the impact of an RA on online consumer behavior and thus provided suggestions for implementing effective systems.

Hao Wang, Qiang Wei, Guoqing Chen [5] has studied From clicking to consideration: A business intelligence approach to estimating consumers' consideration probabilities. It is studied that with rapid advances in e-commerce applications and technologies, finding the chance that a product falls into a consumer's consideration set after being inspected (i.e., consideration probability, CP) becomes an important issue of recommendation services and marketing strategies for both academia and practitioners. This paper proposes a novel business intelligence (BI) approach (namely, the two-step estimation approach, TEA) to estimating CPs with a two-step procedure: one is to introduce partial belongings of consumers to the latent classes with both positive and negative preferences (tastes); the other step is to generate CPs based on the degrees of partial belongings in a weighted probability manner. Experiment results from different online shopping scenarios reveal that TEA is effective and outperforms the traditional latent class model.

C.F. Cheung, F.L Li [6] has studied A quantitative correlation coefficient mining method for business intelligence in small and medium enterprises of trading business. The Business intelligence based on data mining has been one of the popular and indispensable tools for identifying business opportunity in sales and marketing of new products. The traditional data mining methods based on association rules may be inadequate in completely uncovering the hidden patterns of sales based on transaction records. This paper presents a qualitative correlation coefficient mining method which is capable of uncovering hidden patterns of sales and market. Hence, a prototype business intelligence system (BIS) named correlation coefficient sales data mining system (CCSDMS) has been developed and successfully trial implemented in a selected reference site. A series of experiments have been conducted to evaluate the performance of the proposed system. The results generated by the BIS are compared with a well known market available data mining system. The proposed quantitative correlation coefficient mining method is found to possess higher accuracy, better computational effectiveness and higher predictive power. With the new approach, associations for product relations and customer periodic demands are revealed and this can help to leverage organizational marketing capital to enhance quality and speed of promotions as well as awareness of product relations.

2.1 Scope of Investigation

The literature reported the existence of rule based system to study the purchase behavior of the customer however it is not extend further to analyze its result and decide the sales strategy. The online purchase decision making is studied as an effect of unconscious thought from information quality information quantity. It can be extend to develop a conscious thumb rule for further prediction. A quantitative correlation coefficient mining method are used in BI however the BI tools like PENTAHO is found to be more effective in generating the reports for analysis. The small and medium enterprises of trading business are considered so far but it is not extended for a miniature level to a new shop with a single product.

The investigation can be further proposed for the data analysis to study the purchase behavior of the electronics products. This is to study to support the sale strategy for the product. This is followed by the design of the module to study the available past data and current sale to generate the report for analysis using recent BI Tools. This is used to plan the strategy. The correlations are to generate to provide the further prediction. The support system is design and executed for the specific case study relevant to the purchase of electronics product.

3. BI System Architecture

The business intelligence system architecture is designed to study and correlate the stock in hand, sale of products and past trend of sale. This is extended further to develop the graphical results of the analysis using the BI Tool like

Pentaho. The Pentaho has a capability to develop the online reports and their mathematical equations of the trends in graph. These mathematical equation can be used for designing the decision modules. The Structure is shown in Figure 1.

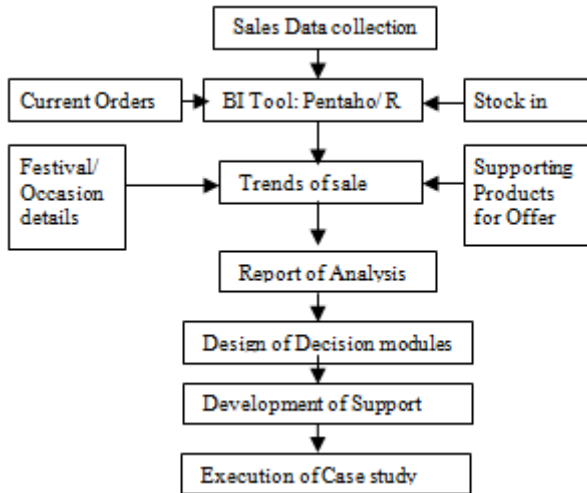


Figure 1: System architecture

The sales data collection is relevant to the sale of the electronics products in specific time duration. The time duration is considered in days. The maximum size of data is collected for the year that is 365 days. The Laptop, Television, Power banks and the Water purifier are four major category of electronics products are consider for the analysis. The three variants in each of the category is selected for the study. The collection of sale data is performed manually by visiting the stipulated electronics outlets. This data is referred for the past sale analysis and its variation. The collected data is reported in Appendix-1. The data shows variation with in 365 days with respect to festivals and special occasions.

3.1 Sales Module for Current Orders

The sale module is designed for four electronics products having three variants each using Prestashop framework. The user interface is open for online shopping and the analysis interface is for the decision support. This module is provided with the customer detail to login for online shopping, It execute the shopping from selection of product to mode of payment. This module also allows the comparison of the products. The analysis interface is designed to control the contents of the shopping portal of the sales module. It consists of dashboard where the profit, Sale and orders variation can be observed graphically. This is also provided with the number of visitors and orders details. The return exchange, incomplete orders and products not in stock is informed on dashboard on the analysis interface. The catalogue is provided to add the number of products along with their details. The details includes product specification, cost, availability and visual images of it. The manufacturer and suppliers details are registered in the catalogue. The order summary details are registered in orders option of this portal. This is helpful to analyse the repetitive orders and to search the key customer, The orders can be reviewed and

update in this module. The orders can be discretized as a day, month and year basis. The Dashboard is shown in Figure 2.

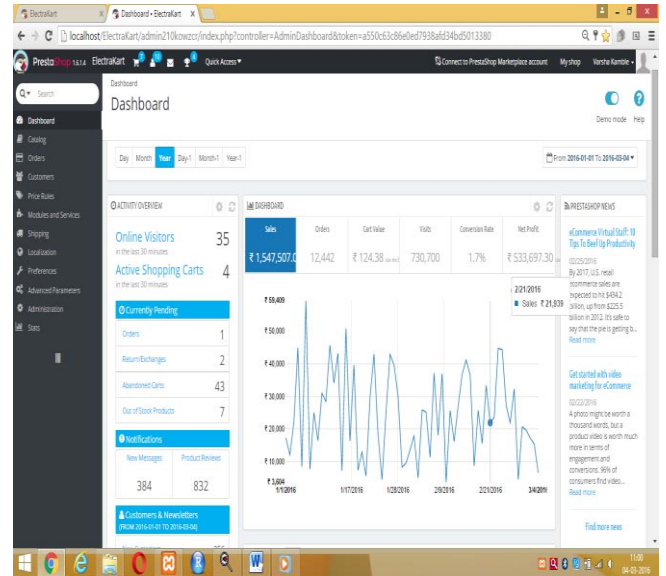


Figure 2: Sales module dashboard

The developed sales module is used to track the current orders. This module has a interface for user and for Analystist. The user can check the products, its specification, reviews essential to place the order. The analysis interface allows to change the product, policies of sale and analysis of profit and sales variation. The past sales data is collected and tabulated manually and can be coupled with the current sale., The execution of the current sale should be design such that the output should update the table of the past sale. This need the separate interface of Business Intelligence tool which convert the sales transaction in to tabulated data along with its graphical explanation. The collective data is stored as a result of sales update and can be studied manually to determine the possible correlation. The correlation is within the sales value against number of days. The obtained correlation is collectively used to design the decision making module. The correlations are called as a heuristic approach whose orientation is developed through the thumb rule. The obtained thumb rule are used only for the prediction and is specific to the particular time bound. The prediction may vary with the variation of data. The obtained heuristic equation can be extend further to apply on a real case study for which the data is available for the specific interval of time. The errors obtained during the execution of the predictions are prepared as a modification and the corresponding modification to the dash board can be performed.

3.2 Pentaho: a BI Tool

Pentaho Report Designer is a sophisticated report creation tool that can be use standalone, or as part of the larger Pentaho Business Analytics distribution. It enables professionals to create highly detailed, print-quality reports based on adequately prepared data from virtually any data source.

Report Designer is one of several ways to create reports with Pentaho software. Through the BA Server's Web-based

Pentaho User Console, we can also use the Interactive Reporting interface, or it can integrate the Pentaho Reporting engine (on which Report Designer is built) into software. This section covers all of the major Report Designer features and functions, from adding a data source to working with conditional formatting and formulas. We can read it cover-to-cover to attain a reasonably comprehensive Report Designer education to use it strictly as a reference to consult when you run into an operational challenge. The main text for your paragraphs should be 10pt font. All body paragraphs (except the beginning of a section/sub-section) should have the first line indented about 3.6 mm (0.14").

The Pentaho database is interfaced with the sales module of the current orders. The data can be collected in the form chart. The chart can be used to study the current trends of the sales variation. This can be generating in the form of report. The report is discretized in different parts namely Header, Footer, and Detail. The recent variation is shown effectively and can be stored as a report for the analysis. The detail of report generation is shown in Figure 3.

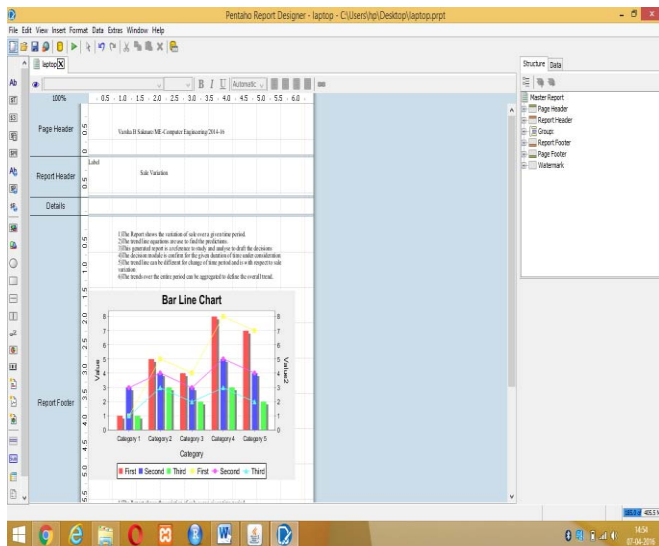


Figure 3: Report generation in Pentaho

3.3 Mathematical prediction using MATLAB

The major problem is to convert the data of Excel sheet in to MATLAB matrix. In This problem, suppose the first column is of days e.g. 1 2 3 4 5 up to 365. In the first column of Excel sheet the days is kept only form row 1 to row 365(No other column in that sheet. The data is vertical (first line of program).It is named with mpa.xlsx. The „a“ is another variable which will read the from mpa.xlsx and save the elements in matrix „a“ which is again vertical column it is reflected in Second line of program. Again x is a matrix where we have to take transpose of a so that the data from mpa.xlsx is available in matrix of MATLAB horizontally with title x. It is reported in third line of the program. It will be helpful to operate in MATLAB Hence following is a syntax to convert mpa.xlsx file in to MATLAB

```
filename='mpa.xlsx';
a=xlsread (filename)
x=a'
```

Similarly the data for other electronics products which includes power bank , purifier and television is saved in .xlsx with file name mpa1 , mpa2 , mpa3 with only on first column and saved in MATLAB with above procedure as a file name of y,m,z,m,m respectively .i.e. in .xlsx it having title mpa1 , mpa2 , mpa3... and in MATLAB y,z,m,.

```
% filename='mpa1.xlsx';
% b=xlsread(filename)
% y=b'
% filenameb='mpa2.xlsx';
% c=xlsread(filenameb)
% z=c'
```

For first trial run the program with only x filename='mpa.xlsx'; a=xlsread(filename) x=a'

The second indicate the product set value in MATLAB with title „m“

```
filenamec='mpa3.xlsx';
d=xlsread(filenamec)
m=d'
```

The third red set from bellow program create the plot and shows the equation with curve fitting. These equations can be used as a mathematical prediction as shown in Figure 4.

```
plot(x,m,'r-d','markersize',04,'linewidth',1)
xlabel('Days')
ylabel('Television')
```

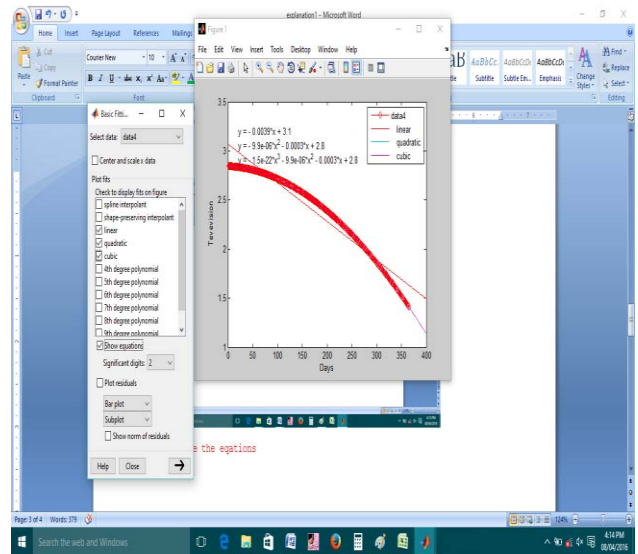


Figure 4: Curve fitting with equation in MATLAB

The above curve fitting offers the mathematical prediction for the sale of products, The obtained reports from the Sales module and pentaho allows the manual entry of data to the excel sheet. The curve shows the corresponding deviation in trend and the equations.

The linear. Quadratic and the cubic trends are obtained along with their equations. These equations are designed as a set to predict the future trends of the sale of electronics products. This will be helpful to control the inventory of the product and minimize the risk to the small scale entrepreneur. The

obtained deviation can be studied to minimize the error and to increase the acceptance of the concept.

4. Conclusion

The Literatures pertaining to the business intelligence is studied and reported for problem definition as a scope of investigation.

The electronics products data for the past year is selected to develop the mathematical predictions. The TV, Laptop, water purifier and the power bank is selected as an electronics products.

The Sales module is developed using prestashop frame work to track the current trends of orders. This trend is registered as a report using Pentaho tool and is used as a reference for analysis and strategy planning.

The MATLAB is used to analyze the sales behavior. It results in finding the curve for the variation along with its equation. The best fitted curve is recommended for the further predictions.

References

- [1] K. Nadia Papamichail , Christopher P. Holland, **2015**, “The effect of prior knowledge and decision-making style on the online purchase decision-making process: A typology of consumer shopping behavior” Decision Support Systems, 51-56
- [2] Hao Wang, Qiang Wei, Guoqing Chen, **2013**, “ From clicking to consideration: A business intelligence approach to estimating consumers' consideration probabilities” Decision Support Systems
- [3] Wingyan Chung, Tzu-Liang (Bill) Tseng, **2012**, “Discovering business intelligence from online product reviews: A rule-induction framework”, Expert Systems with Applications.12-19
- [4] Jie Gao , Cheng Zhang , Ke Wang, Sulin Ba, **2012**, “Understanding online purchase decision making: The effects of unconscious thought, information quality and information quantity”, Decision Support Systems, 87-92
- [5] R. Eric Hostler a, Victoria Y. Yoon etc, **2011**, “Assessing the impact of recommender agents on on-line consumer unplanned purchase behavior Information & Management” , Expert System with application, 15-21
- [6] C.F. Cheung 冫, F.L. Li, “A quantitative correlation coefficient mining method for business intelligence in small and medium enterprises of trading business”, Expert Systems with Applications 39 (2012) 6279–6291

Author Profile



Varsha Saknure is studying in ME Computer Engineering at Zeal Education societys Zeal college of Engineering Narhe, Pune.



Deepak V Jadhav is a assistant professor in computer engineering department of Computer Engineering at Zeal Education societys Zeal college of Engineering Narhe Pune