Review Mining-A Modified Approach for Business Intelligence

Rutuja Jagtap
Students, Department of Computer Engineering
Sinhgad Institute of Technology and Science
Narhe, India

Vishal Kale
Students, Department of Computer Engineering
Sinhgad Institute of Technology and Science
Narhe, India

Gunjan Kshirsagar
Students, Department of Computer Engineering
Sinhgad Institute of Technology and Science
Narhe, India

Shristee Kumari
Students, Department of Computer Engineering
Sinhgad Institute of Technology and Science
Narhe, India

Miss. P.N.Vengurlekar
Assistant Professor, Department of Computer Engineering
Sinhgad Institute of Technology and Science
Narhe, India

Abstract—**In this paper we have explained the concepts and implementation detail of developing a framework which can be used for the purpose of discovering business intelligence from the customer reviews of a product or a service posted online on www.snapdeal.com after the use of that product. Existing work explained the analysis of single product i.e. Samsung S5 on the basis of star ratings given by the customers after the use. Our proposed framework has been designed by collecting the reviews for multiple products of same type. Rather than focusing on star ratings which was done in the existing system, in this proposed system we have considered textual reviews and comments given by the customer since, they are more precise and specific about the quality of the product. In the pre-processing work related to this framework, we have used text mining. The decision generated by the system would be helpful for Customers to know users point of view about product and take purchase decision before buying it as well as it is useful for Business analyst in understanding the product dimensions, attributes and inherent association among them.**

Keywords— Business Intelligence, Data Mining, Product Reviews

I. INTRODUCTION

Now a days, the number of people who prefer online shopping is increasing day-by-day. With the help of online shopping, finding of the desired product becomes pretty easy by viewing multiple products of similar type and by selecting the best among those. Quality is one of the biggest issues when it comes to the online shopping. There are many online shops which sells the similar product. Out of those which is the best one is difficult to find out.

To resolve this problem, we are developing a framework which works on the text mining for multiple products of similar type. The existing system deals with only one product, and also with the star ratings given by the customers. The developing framework analyses the textual comments and reviews given by the customers after the use of those products. By analysing those reviews and comments, the business analysts suggest the company about the quality of the product. Those comments are very useful for improving the quality of the product where it is lagging, and making those best.

This system will work as DSS for the customers, which will help them to make their decision more appropriate by the selection of best product among all the available choices.

Business Intelligence is a process of using the technologies and to make a procedure to look into the historical data of the organization and also to the data collected through various external sources so that information can be retrieved and will help the organization to take better and efficient decision in context to future and will also help to fastened this decision making process. Technologies and procedures used in the BI process provide tools for data analysis and to generate easily understandable business report. A lot of progress has done in the last two decades to develop more advanced tools of reporting and analysing data, and organizations are also very positive to take up these technologies to gain a competitive edge.

Organizations collect data from their routine business operations and also collect data from external sources to get knowledge about their counter parts and to find the answers to queries like- What is marketing policy of the competitors? What is market share of competitors? What are the strategies being adopted by competitors to promote their product etc. And to maintain this information for gaining competitive edge, organizations use different database management and analysis technologies. But these different software technologies are able to store data and to perform data analysis to produce information but to use this information effectively for better decision making, the role of BI comes into picture and

customized software programs are needed to find the answers to above mentioned queries.

DSS is a methodology to support decision making. It can be utilized by a single user or multiple users at multiple users at multiple locations through the web connections. A DSS has been defined as an interactive computer system helping decision makers to combine data and models to solve the problems.

Business intelligence provides a business enterprise with support for better decision making. The development and implementation of BI solutions in organizations is put on the technical parts of the system as most of the efforts is on the design and development and on creating different types of reports.

II. RELATED WORK

T.K. Das et al[2] proposed a system which classify reviews on the basis of star ratings given by the users. Rule based induction approach is used.

Boris Kraychev et al[3] proposed method for information extraction relies on hash function design, which allows comparing and aligning similar trees with a linear computational complexity. The accuracy of the solution was demonstrated over a set of real world web pages.

Anil Sharma et al[4] discussed BI systems and technologies somehow depended on human mathematical Intuitions. Therefore, he suggested that to formulate BI, human mathematical intuitions should not be given the importance rather the importance should be given to the data analysis. Authors further proposed that Materialized Views can be used as a tool to generate business reports with fast query processing in orders to find out real information which further can be used as BI for fast and better decision making process.

Pravesh Kumar Singh et al[5] discussed different methods for data (feature or text) extraction. Also discussed benefits and limitations and one can use these methods according to the situation for feature and text extraction. Based on the survey any one can find the accuracy of different methods in different data set using N-gram feature.

Alexander Pak et al [7] studied how micro blogging can be used for sentiment analysis purposes. He showed how to use Twitter as a corpus for sentiment analysis and opinion mining. Micro blogging tool is used in this paper. Dataset is formed of collected messages from twitter site. Twitter API is used to collect the corpus text posts and formed the dataset of three classes-positive, negative and no sentiments. Naive bayes classification is used for classifying the sentiments.

CHOUDHARY A.K. et al [8] proposed the literature dealing with knowledge discovery and data mining applications in the broad domain of manufacturing with a special emphasis on the type of functions to be performed on data. The major data mining functions to be performed include characterization and description, association, classification, prediction, clustering and evolution analysis.

Herausgegeben von, Robert Remus and Stefan Gindlet al [9] proposed that with sentiment analysis or opinion mining they refer to the task of assigning a sentiment polarity to text documents to determine whether the reviewer expressed a

positive, neutral or negative judgment about a subject. This is an interesting and useful task that has been successfully applied to several different sources of information. For e.g. Movies and Product reviews.

Young Kwark et al [10] discussed about the effect of online product reviews on different players in a channel structure. If two retailers are selling same product manufactured by different manufacturers then how product reviews plays important role to determine the quality of product. The quality of a product refers to the degree of excellence of the product (or of some attribute of the product). For example, the battery life and Wi-Fi connection of a cell phone are about the quality. So Product reviews provide additional information and reduce uncertainties of manufacturers as well as consumers.

David Osimo and Francesco Mureddu et al[11] discussed about outline for discussion upon a new Research Challenge on Opinion Mining and Sentiment Analysis. Opinion mining can be defined as a sub-discipline of computational linguistics that focuses on extracting people's opinion from the web. Sentiment analysis, on the other hand, is about determining about what is the opinion of the writer. Here they focused on the opinion mining applications, tools available in the market for the opinion mining and current researches on the same. This study helps us to know the importance of the sentiment analysis and opinion mining since we are convinced to use customer review mining in our project.

G.Vinodhini et al [12] presents a survey covering the techniques and methods in sentiment analysis and challenges appear in the field. They proposed that an accurate method for predicting sentiments could enable us, to extract opinions from the internet and predict online customer's preferences, which could prove valuable for economic or marketing research. They include different data sources like blogs, micro blogging, data sets and review sites. In the same way our proposed method to analyze the reviews of customers to determine the quality of product referring www.snapdeal.com as data source for online product reviews.
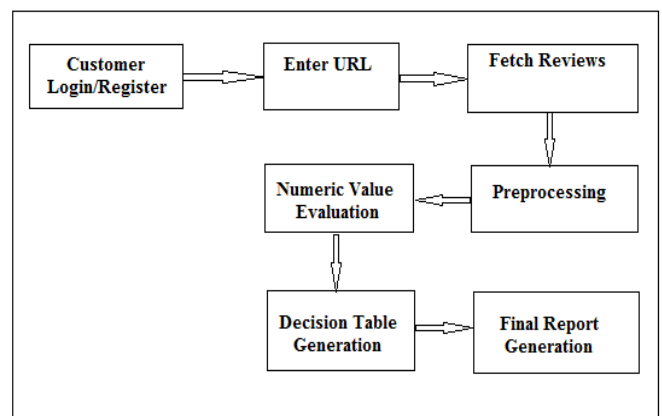
III. ARCHITECTURE
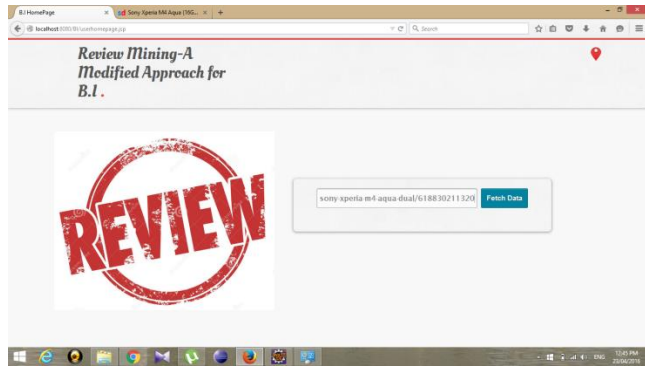

Fig : Architecture

First, confirm that you have

IV. WORKING METHODOLOGIES

A. Customer login/Register

First of all, the customer registers and login to the system for the purpose of authentication.
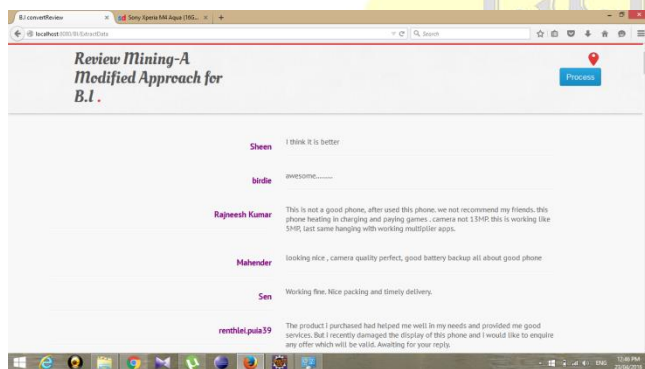
B. Enter URL

Then, we entered the URL of the review page in to the system.

C. Fetch Reviews

After that, with the help of Jsoup library, we fetched the reviews from that page.


D. Pre-processing

In the pre-processing, the concept of text mining is applied. We fetched the comment given by the individual, and decomposed the entire comment into statements, separated by (. /,/...). Then, Each word of the statement get stored in an array and get parsed to determine the adjectives associated with the attributes of the product. After parsing the entire comment, we got all the attributes and adjectives associated with them.
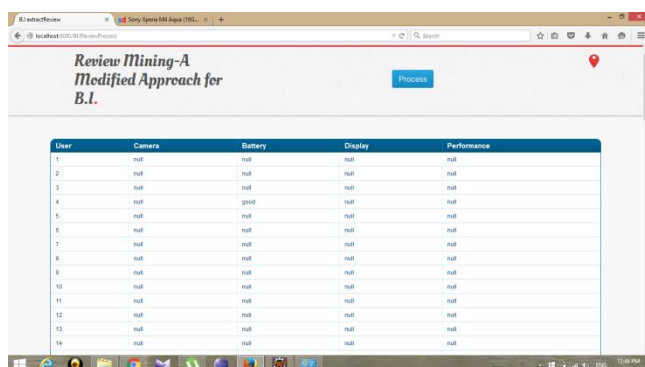

E. Numeric value evaluation

It is difficult to do the calculation on the text, therefore for the ease of calculation we assigned the numeric value to the adjectives associated with attributes.

Example – 5 to excellent or awesome or superb. 4 to better. 3 to good. 2 to ok or average. 1 to bad.
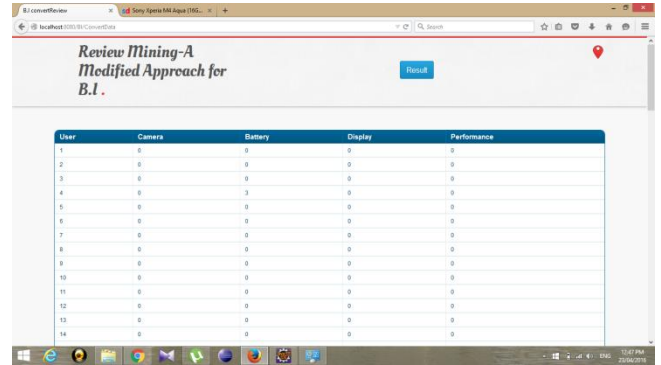
F. Decision table generation

After the numeric value assignment to the adjectives, we calculated the average for each attribute and generated the decision table based on that average.
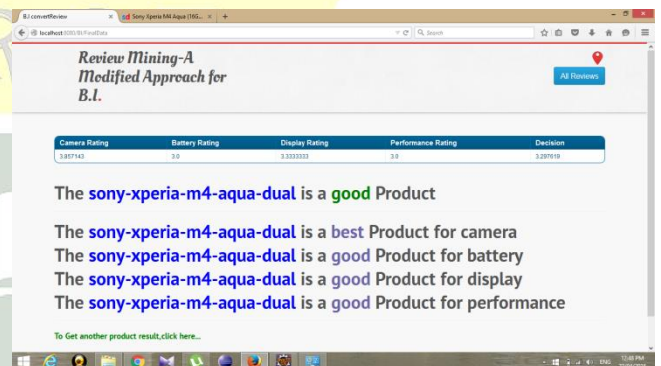
G. Final report generation

In the final report generation procedure, we have calculated the average of produced average of each attribute and gave the final report in the text format about the quality of the product.

V. ALGORITHM

Input:

Url of the review page of any mobile phone from www.snapdeal.com.

Output:

Overall decision about product Mobile phone and its attributes.

Steps for Review Mining

1. Select Attributes.

   (i.e. Features of the phone we need to analyse from the customer reviews.)

2. Save the frequently used adjectives by the customers for those attributes.

(i.e. It will decide the quality of that feature of the mobile phone.)

3. Consider a review paragraph written by customer.

4. Separate the sentences from that paragraph by dot(.), comma(,) or dash(-).

5. Store each sentence in an array.

6. For each sentence recognise the attributes which are selected earlier.

7. Now for extracting reviews.

    For (each word in the sentence in the review database)

    7.1 {

    7.2 If (Selected adjective phrase found)

    7.3 Measure the distance between adjective and attributes

    7.4 If(distance greater than X)

    7.5 Ignore the adjective and attribute

    7.6 Else if(more than one adjective found in the same sentence)

    7.7 Assign nearest adjective phrase to found attribute.

    7.8 }

8. Repeat this procedure for each review.

9. Count all adjectives for each attribute.

10. Assign numeric values to each adjective.

11. Store the result of each review and display in tabular form.

12. For(every attribute){

    12.1 Calculate the average.

    12.2 (Summation of all the assigned numeric values for a single attribute)/(Total no of customer commented on that particular attribute)

    12.3 }

13. Calculate the final average.

    (Summation of average of all the attributes) / (total no. of attributes)

14. Decide categories of average and final average.

15. Enter the features into those categories on the basis of average calculated.

16. Give the overall decision about product

VI. RESULT AND DISCUSSION

In the existing system, they derived the decision rules by using rough set based rule induction LEM2 algorithm. The existing system worked only with the star ratings given for only one product. With the help of those ratings, they concluded about the satisfaction or dissatisfaction about that particular product. From those ratings, it is difficult to determine on which feature user has given the decision about the product.

Review	A1	A2	A3	A4	A5(Decision)
1	5	4	4	5	5
2	5	5	5	5	5
3	1	1	1	1	1
4	5	5	5	5	5
5	2	2	2	1	1
6	2	5	5	3	3

Our proposed system works for textual comments given by the user. On those textual comments, we have applied the text mining and determined the decision of user about that product. We also determined that, on the basic of which attribute user has given the decision about the product. Unlike the existing system, our proposed system works for the multiple products of same type.
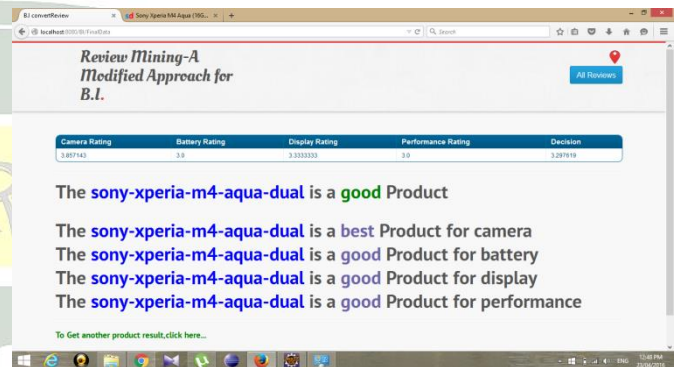
CONCLUSION AND FUTURE WORK

We have developed the system which will mine the reviews automatically and generate a decision table that will help customer to understand the quality of product. It will also help business analysts of that respective company to improve the quality of that product.

Proposed system is developed for multiple products of same type. In future we can develop the system which will mine the reviews for multiple products of multiple types.

ACKNOWLEDGMENT

Every good writing requires the help and support of many people for it to be truly good. We offer our gratitude to Prof. (Miss) P. N. Vengurlekar who encouraged, guided and helped us a lot in this project.

REFERENCES

[1] Bo Pang and Lillian Lee, "Opinion mining and sentiment analysis", Foundations and Trends in Information Retrieval, 2008.

[2] T. K. Das, D. P. Acharjya and M. R. Patra, "Business Intelligence from Online Product Review- A Rough Set Based Rule Induction Approach",2014 International Conference on Contemporary Computing and Informatics(IC3I).

[3] Boris Kraychev, Ivan Koychev,"Computationally Effective Algorithm for Information Extraction and Online Review Mining".

[4]　Anil Sharma, Dr. Manu Sood, "Utilizing Materialized Views to formulate Business Intelligence", 2014 International Conference on Parallel, Distributed and Grid Computing.

[5]　Pravesh Kumar Singh, MohdShahid Husain, "METHODOLOGICAL STUDY OF OPINION MINING AND SENTIMENT ANALYSIS TECHNIQUES", International Journal on Soft Computing (IJSC) 2014.

[6]　Wingyan Chung and Tzu-Liang (Bill) Tseng (2012):"Discovering business intelligence from online product reviews: A rule-induction framework, Expert Systems with applications" 2012.

[7]　Alexander Pak, Patrick Paroubek, "Twitter as a Corpus for sentiment analysis and opinion Mining , Universit´e de Paris-Sud, Laboratoire LIMSI-CNRS, Bˆatiment 508,F-91405 OrsayCedex, France."

[8]　CHOUDHARY A.K., HARDING J.A. and TIWARI, M.K., 2009, "Data mining in manufacturing: a review based on the kind of knowledge", Journal of Intelligent Manufacturing, 20 (5), pp. 501 – 521.

[9]　Herausgegeben von, Robert Remus and Stefan Gindl, "Practice and Theory ofOpinion Mining andSentiment Analysis", Journal for Language Technologyand Computational Linguistics.

[10]　Young Kwark ,Jianqing Chen, SrinivasanRaghunathan, "Online Product Reviews: Implications for Retailers and Competing Manufacturers" , Information Systems Research 2014.

[11]　David Osimo and Francesco Mureddu, "Research Challenge on Opinion Mining and Sentiment Analysis", http://www2.research.att.com/~volinsky/DataMining/Papers 2008.

[12]　G.Vinodhini, RM.Chandrasekaran, "Sentiment Analysis and Opinion Mining: A Survey", International Journal of Advanced Research in Computer Science and Software Engineering , 2012

[13]　PrernaMahajan, RekhaKandwal, Ritu Vijay, "Rough Set Approach in Machine Learning: A Review",International Journal of Computer Applications, 2012

[14]　Sofien GHARBI,HamzaBELLAKHDAR , Salah Eddine MRABET, "Project based learning in business intelligence with intervention of companies", IEEE Global Engineering Education Conference (EDUCON), 2015.