



Journal Homepage: -[www.journalijar.com](http://www.journalijar.com)  
**INTERNATIONAL JOURNAL OF  
 ADVANCED RESEARCH (IJAR)**

Article DOI:10.21474/IJAR01/7958  
 DOI URL: <http://dx.doi.org/10.21474/IJAR01/7958>



## RESEARCH ARTICLE

### DATA ANALYSIS ON CUSTOMER REVIEW.

Akshat shah, karan shah, hridayesh shah and dhruv shah.

#### Manuscript Info

##### Manuscript History

Received: 17 August 2018

Final Accepted: 19 September 2018

Published: October 2018

#### Abstract

Now-a-days, e-commerce websites have million of products ,each having hundreds of reviews. While buying any product, a customer spends a major time in reading its reviews and getting feedback from the past customers. This is an integral part of the ecommerce experience since the customers takes its decision solely based on others' opinions. The problem is that majority of reviews are not proper which compromise the genuine buyer's ratings. These reviews can mislead the customer to degraded shopping experience.

This is where the customer review analytics comes into picture. We examine the reviews on online shopping like amazon by various data mining methods. We analyze the available reviews of the product till date to give the buyer better and proper rating at a time of purchasing the product. This will also give the feature specific ratings which also help customer to analyze the two different products. We also analyze the reviews for identifying the person's perspective towards the product.

For the development of this project, .Net Framework architecture, SQL server and some data mining algorithms for processing and classification of data will be used. The data in the link provided by the user, is scraped, pre-processed and classified. The scraping will be done with the help of web scraper. Once the scraped data is obtained, it will be preprocessed and classified with the help of data mining methods. The result of the classification process generated is then provided to user. This result will then tell the user about the ratings of the product based on reviews.

The algorithm used in the project will provide accuracy on the amount of precise and correct data. This would help customers in confidently buying a product. Moreover, e-commerce platforms would be benefited by increased customer retention. Since there is shift from traditional buying to the web based purchasing our solution fits rightly in the market.

Copy Right, IJAR, 2018,. All rights reserved.

#### Introduction:-

The growth of E-commerce has led to the invention of several websites that market and sells products as well as allows users to post reviews.

It is typical for an online buyer to refer to these reviews before making a buying decision. Hence, automatic

summarization of users' reviews has a great commercial significance.

However, since the product reviews are written by non experts in an unstructured, natural language text, the task of summarizing them is challenging.

This Project presents a semi supervised approach for mining online user reviews to generate comparative feature-based statistical summaries that can guide a user in making an online purchase.

### **Problem Definition:**

#### **Phase I**

##### **Planning:**

Planning will involve a comprehensive suite of project plans which will set out a clear project roadmap ahead.

We will initially identify the need of analyzing the reviews available online.

Resources required to develop the project will be identified i.e. requirement gathering will be done. On the basis of the requirement of the project, the cost and total time required to complete the project will be estimated.

##### **Analysis:**

1. This phase will specifically address establishing a baseline and a way to track the requirements through the rest of the lifecycle.
2. The data regarding customer reviews will be gathered from various e-commerce websites.
3. We will be theoretically comparing various classification algorithms and the algorithm which we will be using for classification will be decided.

##### **Design:**

1. After gathering the requirements, the blueprint for the data analysis system will be designed.
2. With the help of the blueprint will be able to identify errors that may arise in the final system.

##### **Coding:**

1. We will be using HTML, CSS, javascript for creating a user-friendly website.
2. We will be using .net framework for fully implementation of the application.
3. We will be using SQL server with BI tools for analyzing the data available with us.

##### **Implementation:**

On the basis of the product rating required by the user, the data will be scraped and will be stored in an object which will be preprocessed. This data will be stored in csv. On this csv file, classification will be done and the result will be provided in the form of ratings and graphical representation of the reviews over the years.

#### **Phase II**

##### **Testing:**

1. With respect to UI, initially the browser compatibility of the UI will be checked, also the correctness of the responsive web design will be tested.
2. Test cases will be determined to test whether the data scraped is correct or not and the scraped data is correctly preprocessed or not.
3. Various test cases will be determined in order to check the accuracy of the result and also limitations of the system.
4. The result of these test cases will be used for further improvements in our system.

##### **Deployment:**

Once the testing is done, the system is now ready for deployment. Deployment activities will include the release, activation, adaptation, updates, version tracking of the system etc. The user will now be able to access the system, where the user can get analysis of the customer reviews

#### **Methodology Used**

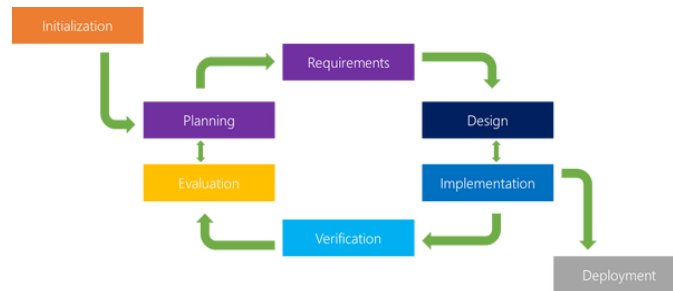
For developing a project or software it is important to know the flow or which type of software development lifecycle model we are going to use. Out of various SDLC model we need to choose the model which will be perfect

for our project and therefore we are using Iterative model.

As the name suggest Iterative model is basically an iterative process where the input is processed through various phases and if it doesn't satisfy it is repeated until it becomes satisfactory.

The model consists of five phases namely:

1. Planning.
2. Analysis and Design.
3. Implementation.
4. Testing.
5. Deployment.



**Fig 2.1: Methodology**

### **Planning:**

For developing any software or developing any project, planning plays a crucial role, as without planning a project won't reach its maximum level. The planning phase deals with requirements of customer, understanding what the customer wants and making it available for the customer. It is also important to understand the purpose, scope and goal of project.

For developing the system giving analysis on customer reviews , initially we have understood in what way does the customer want the results, and thereby planned the phases of our project accordingly.

### **Analysis and Design:**

Once the requirements and purpose of the project is understood it is important to analyze and work on the design of the project. Depending upon the requirement, the design of the project needs to be visualized. The design of the project should satisfy basic criteria like easy access to users, low cost, etc. To decide the work flow of system various diagrams like DFD, use case, flow chart, etc. need to be created. In this phase we have designed the layout for our project along with DFD UML and flow chart which help in providing better understanding of our system.

### **Implementation:**

Once the design of system is ready the part of implementing it, i.e. coding begins. The implementation of system is to be carried out in a systematic manner. First the UI (User Interface) is to be created with the help of web development languages namely HTML, CSS, JAVASCRIPT, etc. then the server-side processing is to be done. In our project the server-side processing is carried out with the help of .NET framework and MYSQL server which is used for scraping and classification purpose and lastly the integration of both. After completing implementation, the system is partially ready.

### **Testing:**

Testing plays an important role in any SDLC because without testing your product you cannot deploy it for user. Testing needs to be carried out in two ways black box and white box where the team can understand both the scenarios and can solve the upcoming problems in order to improve the system.

### **Deployment:**

Once all the testing is done, the system is ready to be deployed and to be used by user. The web portal now can be hosted and the working of the system is to be monitored. Feedbacks from the users will be beneficial for improvement.

**Feasibility Study**

The very first phase in any system development life cycle is preliminary investigation. The feasibility study is a major part of this phase. A measure of how beneficial or practical the development of any information system would be to the organization is the feasibility study.

The feasibility of the development software can be studied in terms of the following aspects:

1. Operational Feasibility.
2. Technical Feasibility.
3. Economical feasibility.
4. Motivational Feasibility.
5. Legal Feasibility

**Operational feasibility :**

The site will reduce the time consumed to maintain manual records and is not tiresome and cumbersome to maintain the records. Hence operational feasibility is assured.

**Technical feasibility :**

The following technical requirements are necessary for the implementation of project:

1. At least 166 MHz Pentium Processor or Intel compatible processor.
2. At least 512 MB RAM.
3. 14.4 kbps or higher modem.
4. A mouse or other pointing device.
5. At least 50 GB free hard disk space.
6. Chrome or any other compatible browser

**ECONOMICAL FEASIBILITY :**

Once the hardware and software requirements get fulfilled, there is no need for the user of our system to spend for any additional overhead.

For the user, the web site will be economically feasible in the following aspects:

1. The web site will reduce a lot of paper work. Hence the cost will be reduced.
2. Our web site will reduce the time that is wasted in manual processes.
3. The storage and handling problems of the registers will be solved.

**Legal feasibility :**

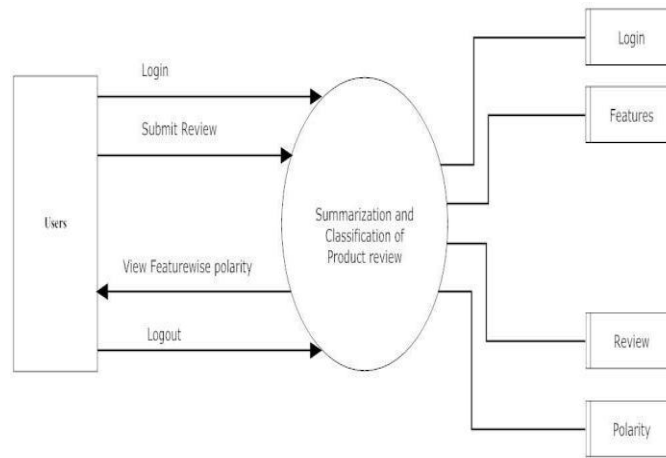
The licensed copy of the required software is quite cheap and easy to get. So from legal point of view the proposed system is legally feasible.

**Requirements**

1. Web browser
2. Windows Operating System
3. Windows XP, Windows 7,8
4. My SQL server
5. Visual Studio 10

## Design and Implementation

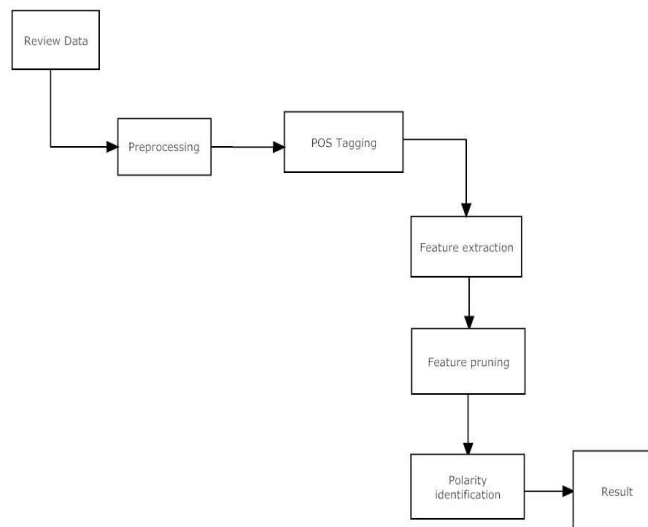
### DFD:



**Fig 5.1:-Data Flow Diagram**

Shown above is the Data Flow Diagram where the overall flow of data from user to database. As we can see user can login to the system and add his/her review on a product and can even see the overall rating of the product as well.

### Flowchart:



**Fig 5.2:-Flow Chart**

## Results and Discussion:-

### Output:

1. Our project will give the output as the ratings on the reviews gathered from different websites about specific product.
2. We will also get the analysis of each and every feature include in the review sentence.
3. It will also analyze the reviews with the emoticons include in it and will give the result and output according to the reviews .
4. The rating will also be given the feature specific so it will be giving us ratings not only based on the reviews but also features of products that consumer is going to buy.

### Conclusion:-

Classifying and summarizing reviews of customers has several interesting and commercially significant applications. However, this task is much more difficult than classifying regular reviews sentences and requires intensive Preprocessing. The success of the reviewing the sentiments behind the reviews is task which mainly dependent on the efficiency and sophistication of the Preprocessing and extraction of sentiments of customers . We have proved that the method which we used to analyze the data is indeed helpful to the customer

The results obtained in last through the analysis of customer reviews gives the ratings by analyzing each and every features in sentence of review available and classify their orientation with acceptable accuracy. This enables reliable review rating which has several commercially important applications like we can use it in e-market for purchasing any specific product.

As the analysis of customer reviews , we can use it for giving the proper representation of the product with respect to the customers point of view. This will improve the efficiency of the system and reliability of the customers. Also it will reduce the fake reviews use to manipulate the customer, which help customer to obtain pure analysis through the system.

This system will efficiently increases the customers reliability with result and data sufficiency to improves prospects toward the product he/she want to purchase.

### Future Scope

In the future, we want to perform data mining on larger and more varied customer reviews data sets. We would also like to extend our work to fuzzy opinion classification to support fuzzy user querying. We are also intending to make the reviews as much as useful to the customer who wants to buy any specific product. The classification technique proposed for the classifying the impact of the reviews on product. We are also intending to make the system more efficient and more faster to satisfy the customer.

### References:-

#### Journal References:

1. D. Mali, M. Abhyankar, P. Bhavarathi , K. Gaidhar, M.Bangre “Sentiment Analysis Of Product Reviews For Ecommerce Recommendation” in IJMAS, vol 2 issue 1, pp 127-131, Jan 2016
2. Wararat Songpan “The Analysis and Prediction of Customer Review Rating Using Opinion Mining” in IEEE sera, pp 71-77, June 2017
3. Shengsheng Xiao , Chih-Ping Wei , Ming Dong “Crowd intelligence: Analyzing online product reviews for preference measurement” in Elsevier, at national Taiwan university, pp 169-182, October 2016
4. Chong, Alain Yee Loong and Ch'ng, Eugene and Liu, Martin J. and Li, Boying “Predicting consumer product demands via Big Data: the roles of online promotional marketing and online reviews” in IJPR, pp 5142-5156, at Nottingham university, June 2017
5. Patrick De Pelsmacker , Sophie van Tilburg , Christian Holthofb “ Digital marketing strategies, online reviews and hotel performance ” in Elsevier, at Artesis Plantijn University, pp 47-55, Jan 2018