

Best Customer Services among the E-Commerce Websites – A Predictive Analysis

M. Emerentia, N. Yuvaraj,

PG Scholar, KPR Institute of Engineering and technology, Coimbatore. INDIA

rinisoundhar@gmail.com

Assistant Professor, KPR Institute of Engineering and technology, Coimbatore. INDIA

yuvaitksr@gmail.com

ABSTRACT :

Online shopping websites have become popular as they allow customers to purchase the products easily from home. These websites requests the customers to rate the quality of the products. To enhance the quality of the products and services these reviews provides different features of the products. In order to buy a product a customer has to go through large number of reviews, which is quite hard for the customers as well as the marketers to maintain, keep track and understand the views of the customers for the products. In this project I have designed a consolidated website where I have the comparison data about the e-commerce website services such as Price, cash on delivery, Product replacement etc., and gave the predictive result about the shopping website which gives the best customer services. This can ease the purchasing time of the customer instead of comparing the shopping websites to review about the product and finding the best e-commerce website to purchase the product.

1. INTRODUCTION:

Sentiment Analysis or Opinion mining is one of the emerging research fields in Data Analytics. It is one of the popular techniques which is used for analyzing and summarizing the customer's reviews about the products and the services. Some of the online shopping websites such as Amazon, Flipkart, Ebay, Snapdeal, etc contains the product reviews which enable the customers to go through the reviews while purchasing the product. Sentiments are extracted from the feedbacks of the products given by the users from the shopping websites. Extraction of opinion includes identification of opinion holder, subject of the review and the responses such as positive or negative.

This paper explains about the sentiment analysis techniques used for predicting the best customer services among three shopping websites such as Amazon, Flipkart and Snapdeal using the machine learning algorithm. First a Sample website is developed and a stream of data such as the delivery time , price , COD etc., is gathered from the shopping websites. We use the Support Vector Algorithm for the classification of reviews. Based on the average result of the positive and negative reviews, the best customer service provided by the shopping websites can be predicted.

2. LITERATURE SURVEY:

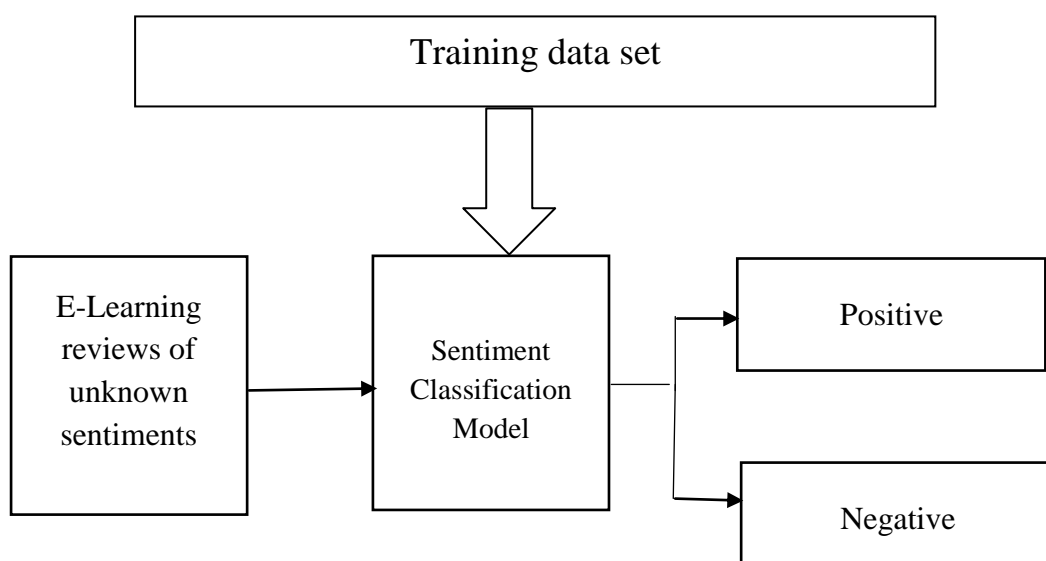
SVM classifier fits well for the Sentiment analysis providing an accuracy of about more than 80% since they have the high-dimensional input space and Document vector space. SVM uses a method known as over fitting protection which doesnot depends on the total number of features which increases the ability to handle large number of features. The only disadvantage of SVM is suppose if any of the categorical or missing values is found it needs to be preprocessed. Low processing memory requirement and less time of

execution is the main advantage of Naïve Baye’s classifier .It is advisable to use this algorithm when the training factor seems to be a crucial factor for in the system. The main disadvantage of this classifier is that the assumptions of the attributes being independent which may not be valid. When comparing the Naïve Bayes

and NN algorithms, the accuracy was extremely significant when the training dataset is small as 50, 150, and 200. NN algorithm has the highest computation cost because it is necessary to compute the distance of each instance of the query for all the samples of the training dataset and also it has the poor run time performance when the dataset is very large. NN algorithm is not suitable for distance based learning like SVM because it is not clear to know about which attributes will match the particular distance to provide better results. Finally on comparing the three algorithms SVM gives the best accuracy rate of about more than 80% for large training datasets.

3. OPINION MINING PROCESS:

Opinion mining or sentiment analysis has attracted great interest in recent years. One of the most promising applications is analysis of opinions in social networks. People write their opinions in forums, micro blogging or review websites. The data is very useful for business companies, governments, and individuals, who want to track automatically attitudes and feelings in those sites. Namely, a lot of data that contains much useful information, so it can be analyzed automatically. On an instance, if a customer wants to buy a product he usually searches the Web trying to find opinions of other customers or reviewers about this product. These kinds of reviews affect customer’s decision. Sentiment Analysis or Opinion mining is one of the emerging research fields in Data Analytics. It is one of the popular techniques which is used for analyzing and summarizing the customer’s reviews about the products and the services. Some of the online shopping websites such as Amazon, Flipkart, Snapdeal, etc contains the product reviews which enables the customers to go through the reviews while purchasing the product. Sentiments are extracted from the feedbacks of the products given by the users from the shopping websites. Extraction of opinion includes identification of opinion holder, subject of the review and the responses such as positive or negative. Opinion mining task can be transformed into classification task, so machine learning techniques can be used for opinion mining.



Sentiment Classification Mechanism

4. IMPLEMENTATION:

The implementation process consists of the following methods:

- Collection of the data
- Classification Process
- Architecture Evaluation
- Finding the Average

4.1 COLLECTION OF THE DATA:

- Data is collected from the shopping websites such as Flipkart, Snapdeal, and Amazon.
- Collection of the data is mainly focused on the customer services of these shopping websites that includes Cash on delivery, Delivery time, product replacement, etc.
- Using these data a comparative database is generated which makes the customer to easily review the services offered by various e-commerce websites.

Details	Flipkart	Snapdeal	Amazon
PRICE (in Rs)	18,499	20,498	18,000
COD	Available	Available	Available
DISCOUNT	Nil	28%	18%
DELIVERY TIME	4-5 DAYS	2-3 DAYS	2-3 DAYS
PRODUCT REPLACEMENT	30 DAYS	15 DAYS	20 DAYS
SHIPPING CHARGE	NIL	NIL	Nil
DELIVERY TYPE	STANDARD	FAST	STANDARD
AVAILABILITY	AVAILABLE	AVAILABLE	AVAILABLE
OFFERS	NILL	Airtel 4G simcard	NIL
EMI PAYMENT(in Rs/month)	990	1,260	1607

Cancel

4.2. CLASSIFICATION PROCESS:

To classify data into predefined classes is the goal of text classification ie, positive and negative classes. Text classification comes under the supervised learning problem.

Transforming document which is in string format into format suitable for learning algorithm and classification task is the main step in the text classification..

4.2.1 SVM LEARNING ALGORITHM FOR TEXT CATEGORIZATION:

Support vector machine has the predefined input and the output format where the input is a vector space and the output will be either 0 for negative or 1 for positive [9][5]. Using the pre-processing techniques the documents which are in original format is converted into machine learning input. The best classification of data techniques is supported by SVM [6]. SVM is one of the learning algorithms which provides best learning algorithm for text categorization. For 1000 features SVM provides High Dimensional Input Space by applying over fitting protection that does not depends on the number of features, which provides high ability to handle the features in large numbers [9]. SVM provides sparse Document Vector Space since the document vectors contains about very few non-zero elements [7]. SVM's task is to generalize the mapping of input-output.

4.2.2. SVM BENEFITS:

1. High Dimension Input Space: In text classification we have to deal with many features (may be more than 1000). Since the over fitting protection in SVM which does not depend on number of features which increases the ability to handle large number of features[1][2][9].

2. Document Vector Space: The document vectors consists of only a few non-zero elements, despite the high dimensionality of the representation[1][2][9].

4.2.3 SVM PERFORMANCE & EVALUATION:

For the text categorization, the input dataset will be the documents and the generated output will be the class of the respective documents. For example, In spam filtering the input will be the e-mail and the output will be 0's and 1's which indicates the negative as Spam and the positive as Not-Spam. The basic objective behind the SVM classification is to find out the hyper plane [3] that has maximum margin which separates the document vector of one class from the document vector of another class with maximum margin. SVM constructs a set of hyper planes or a set of hyper-planes with a high or infinite-dimensional space. A hyperplane having largest distance to the nearest training data points provides the good separation i.e., larger the margin the lower will be the generalization error of the classifier.

For a training set D , a set n points can be written as,

$$D = \{(x_i, c_i) \mid x_i \in \mathbb{R}^p, c_i \in \{-1, 1\}\}_{i=1}^n \text{ -----1}$$

Where x_i is a p -dimensional real vector. Then find the maximum-margin hyper plane for the points $c_i = 1$ and $c_i = -1$ where $\{1, -1\}$ corresponds to positive or negative data [11], which can be given by,

$$w \cdot x - b = 1 \text{ and } w \cdot x - b = -1 \text{ -----2}$$

the distance between the planes is given by, $\frac{b}{\|w\|}$ in which $\|w\|$ has to be minimized .

Using Lagrange's multipliers the optimization problem can be expressed as ,

$$\min_{w, b} \max_{\alpha} \left\{ \frac{1}{2} \|w\|^2 - \sum_{i=1}^n \alpha_i [C_i (w \cdot x_i - b) - 1] \right\} \text{ -----3}$$

From the equation 2 & 3, the α_i should be greater than zero because they are the document vectors corresponding to w . Using this method SVM determines the hyperplane's location corresponding to the w 's side i.e., $\{1, -1\}$ [8]. SVM gave the accuracy of about 82.9% and also it provides the low dataset dimensionality.

4.2.4 POSSIBLE OUTCOMES AND RESULT:

SVM has defined input and output format. Input is a vector space and output is zero or one (positive/negative). Number of words increase if increase the number of documents. Practically infinitely many features can be constructed which can enhance classification accuracy. Task of SVM is to learn and generalize the input-output mapping. In case of text categorization input is set of documents and output is their respective class. Consider spam filter as example input is an email and output is 0 or 1 (either spam or no spam)[9].

4.2.5 Why SVM ?

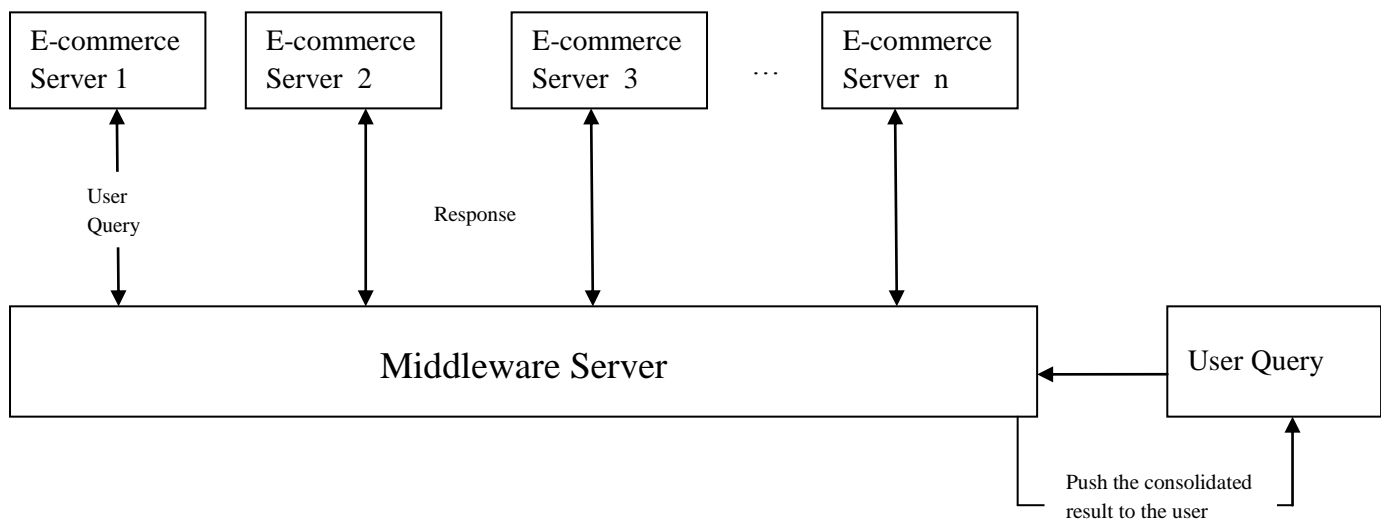
It is found that SVM classifier fits well for the Sentiment analysis providing an accuracy of about more than 80% since they have the high-dimensional input space and Document vector space[1][2]. SVM uses a method known as over fitting protection that increases the ability to handle large number of features since it does not depend on the total number of features. The only disadvantage of SVM is suppose if any of the categorical or missing values is found it needs to be preprocessed.

4.3. ARCHITECTURE:

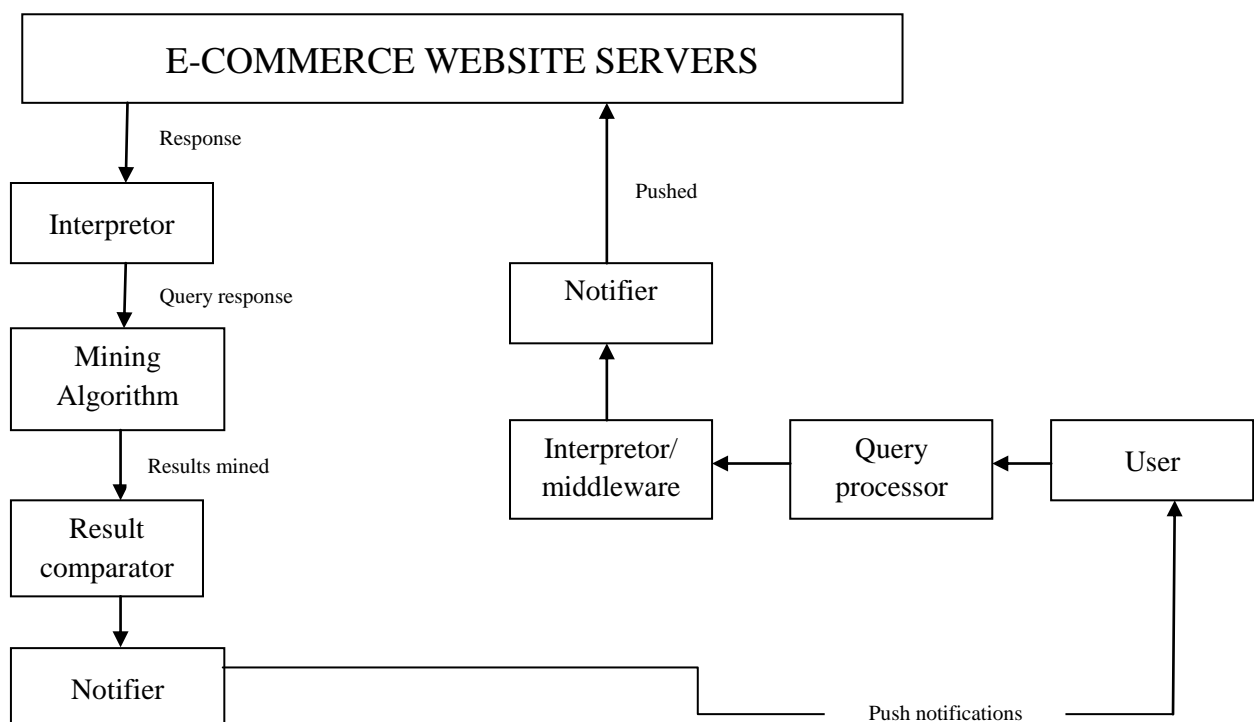
The architecture consists of two types namely:

- Conceptual Architecture
- Middleware server Architecture

4.3.1 CONCEPTUAL ARCHITECTURE:

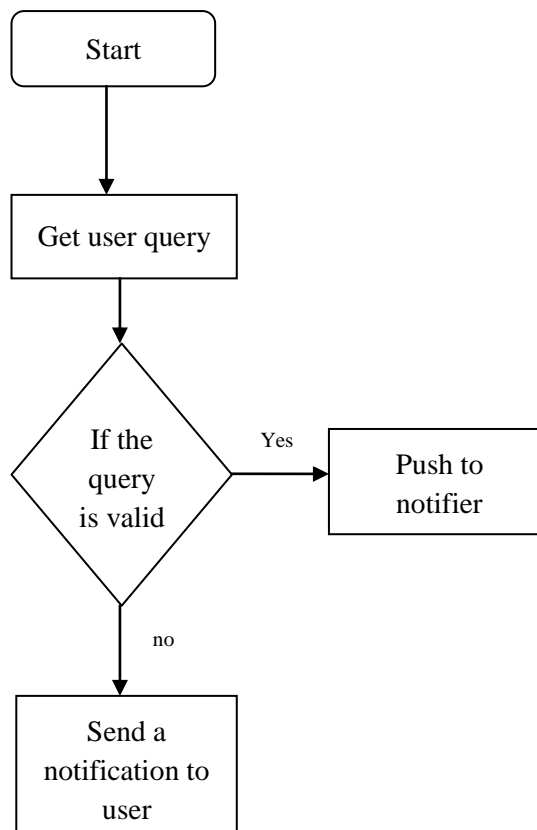


4.3.2 MIDDLEWARE SERVER ARCHITECTURE:



4.3.2.1 Description about the architecture:

The architecture shows how the user query is processed by the server. The query which is given by the user for example (iphone 5s) is first processed into the middleware server where the query is analyzed and processed by the query processor. The query which is processed is then sent to the interpreter which examines whether it is a valid query to be sent to the server. After this process the query is transferred to the notifier where it will be notified to the e-commerce website servers. The middleware server, after getting the response from the e-commerce servers it will send the response query to the responding interpreter where the query is validated again. Once after the completion of the validation process by the interpreter the mining process will be executed. The mining process will be executed by the classification algorithm. The algorithm evaluates the database and provides the mined results to the result comparator where the results are compared and consolidated according to the data given by the servers. The consolidated data is now sent to the notifier where the notification about the processed query will be pushed back to the user.



Flow diagram for the Interpreter Process

4.4 FINDING THE AVERAGE:

From the result obtained from the classification process, the following process has to be Carried out:

- ❖ The average number of positive reviews has to be calculated for the three shopping websites.

- ❖ The average number of negative reviews has to be calculated for the three shopping websites.
- ❖ Compare the results of the positive and negative reviews.
- ❖ The shopping website having the more number of positive reviews is predicted as the best customer service providing website.

5. RESULT & CONCLUSION:

The wide varieties of sentiment analysis applications are there in various systems that include classifying and summarizing reviews. We here used the Support Vector machine algorithm to find the polarity of sentiments and classified them accordingly. The sentiment analysis for the products review will help the customer to choose the best product[10]. It will help to identify the quality of product. Also it will help the developer or company to remove the disadvantages of their product or services and re-design them according to customer's need. It is a cost effective approach for capturing the feedbacks of consumers[10].

About Authors:

M.Emerentia received her Bachelor's Degree in Information Technology. Now pursuing M.E degree in Computer Science and Engineering in KPR Institute of Engineering and Technology, Arasur, Coimbatore.

Mr. N. Yuvaraj, Assistant Professor(Sr.G) in Department of Computer Science and Engineering, KPR Institute of Engineering and Technology, Coimbatore. He has completed his Bachelor of Technology in Information and Technology and Master of Engineering in Software Engineering. He has seven years of teaching and two years of industry experience. He has published over 10 technical papers in conferences and journals. His area of research includes Data mining, data analytics and information retrieval.

REFERENCES:

1. Pravesh Kumar Singh, Mohd Shahid Husain ,Methodological Study of Opinion mining and Sentiment Analysis Techniques, International Journal on Soft Computing , Vol no , 5, 2014
2. Amit Gupte, Sourabh Joshi, Pratik Gadgul, Akshay Kadam, Comparative Study of Classification Algorithms used in Sentiment Analysis , International Journal of Computer Science and Information Technologies, Vol. 5 , 2014.
3. Safa Ben Hamoud, Jalel Akaich, Social Networks' Text Mining for Sentiment Classification: The case of Facebook' statuses updates in the "Arabic Spring" Era, International Journal of Application or Innovation in Engineering & Management (IJAEM), Volume 2, Issue 5, 2013
4. <http://people.revoledu.com/kardi/tutorial/KNN/Strength%20and%20Weakness.htm>
5. Rohini S. Rahate, Emmanuel M, Feature Selection for Sentiment Analysis by using SVM, International Journal of Computer Applications, Volume 84, Issue No 5, December 2013
6. Jayashri Khairnar, Mayura Kinikar , Machine Learning Algorithms for Opinion Mining and Sentiment Classification, International Journal of Scientific and Research Publications, Volume 3, Issue 6, 2013.
7. Ankush Sharma , Aakanksha , A Comparative Study of Sentiments Analysis using Rule Based and Support Vector Machines, International Journal of Advanced Research in Computer and Communication Engineering, Vol. 3, Issue 3, 2014.
8. S.ChandraKala, C.Sindhu, Opinion Mining And Sentiment Classification:A survey,CTACT Journal On Soft Computing,vol.No, Issue.0, October 2012.

9. Gaurangi patil,Varsha Galande,Vedant Kekan,Kalpana Dange,Sentiment Analysis Using Support Vector Machine,International Journal of Innovative Research in Computer and Communication Engineering,Vol.2,Issue 1,January 2014.
10. Dhanashri Chafale, Amit Pimpalkar,Corpora for Sentiment Analysis of Product reviews using Emotions Wheel and Fuzzy logic, International Journal of Engineering Research Online,Vol.3.,S2.2015.
11. Bo Pang and Lillian Lee, Shivakumar Vaithyanathan, Thumbs up? Sentiment Classification using Machine Learning Techniques, Proceedings of EMNLP, 2012