

# Monitoring the Best product from Consumers opinion forums based on their Budget using Data Mining Technique

Jinesh Melvin Y I<sup>#1</sup>, Teena James<sup>\*2</sup>, Suresh Babu<sup>#3</sup>

<sup>#</sup>Assistant Professor, Computer and Information Technology & Mumbai University, India

**Abstract** — In today's market scenario consumers reviews play an important role in deciding whether a product should be purchased or not. In most of the product analyzing sales report, they check the customer's reviews, rating, company, product models and so on. In the previous existing module they analysed the products from customers rating and reviews, but there was no proper solution for customers to buy products. In this paper we decide to monitor the best product from consumers reviews based on their budget. Here we collect all products online reviews from different websites also oral reviews from different customers. We introduce this module for customers goodness and their satisfaction, aiming both low budget buyers and high budget buyers. Our reviewing mechanism econometric analysis deals with data mining techniques. This module will give a clear idea about the product for different budgeters. Our results can have several implications for the market design of online opinion forums.

**Keywords** — Data Mining, Review Analytics, Internet Bot, Weka.

## I. INTRODUCTION

In day to day life, people are forced to buy more product for their lifestyle and day by day more different products get introduced in the market. Now a day's people face more troubles because of quality-less product, High budget with low quality and duplicate products. People get confused to purchase the products from online and offline. To overcome these problems many of the enterprises get started to analyse the product from customer's opinion.

Many people give the priority to on-line shopping over the Traditional Shopping, for that they are using smart devices such as tablets, mobile phones, laptop and desktop to access E-Commerce websites through the Internet. In addition, they want to get their desired product in the best price. In order to get their desired product in their budget they would like to survey in different E-commerce sites. But all sites and technology are not satisfied the users demand due to restriction. Our product analysis system overcomes these issues and it supports for different customers decisions.

This system will give more influence to both online and offline consumers. It will help to decide the best product for different budgeters, hereby using this technology user can get their desired product in minimum price apart from these features this technology monitors best to buy product to user over the numbers of E-commerce website. Normally if any companies or industries before launched their new product in the market they give advertisement in any E-commerce website. From that customers start to post their reviews and rating. Once the product get launched in the market, people try to survive the product with the help of different websites. Our aim is to analyse the products reviews and its budget from online and offline. Most of the people would like to buy cheap and best product but it's not for all. In this paper we quoted the data mining technique and its working principles.

## II. LITERATURE REVIEW

Data mining techniques were mentioned for monitoring the best reviews. In paper [1] focus on fetching the URL's from different websites for analyzing the products. The Real Time Product analysis using data mining enables the buyers to compare products from different E-commerce websites, thus facilitating them to purchase the product at the best deal. The main elements of this technology are Web crawler and Web scraping. The main purpose of the crawler is to crawl different E-commerce websites and to fetch the URLs of the products from these websites. Every E-commerce website can be considered as a graph consisting of several nodes (Links or URLs). The crawler must traverse to all these nodes and fetch these nodes. Once it has fetched the node, that node must be kept in a set of visited nodes so that no two same URLs are fetched. Threads that are created in the thread-pool must be limited so that they do not eat up the entire memory. And each thread that's been started has to be terminated. The Coordinating thread distributes the crawl job to the processing threads. These processing threads fetches the URLs and returns to the Coordinating threads. Thus, the fetched URLs that we have in the set visited nodes are given to the scraper for scraping purpose. Web scraping can be defined as a process of extracting HTML data from the URLs and then using this data

for personal purposes. Once we fetched the URLs with us then the job is to get the information that is abstracted within the URL. For example the mentioned URL  
<http://www.snapdeal.com/product/htc-desire-526-g/681509924248>.

Data mining of sales data gives frequent pattern of same type of product sales. It is like market basket analysis within same type of products and by representing this data graphically we get easily and quickly view type of customer. So we can target the customers which will increase sales and profit of organization. Based on the Association Rule, sales analysis will produce a result for product rating. We take the rating from the user and also generate rating from system by taking this further we calculate the average mathematical rating for a particular product. We also see which product sales more in the particular categories, by using above both behavior of customer and sales report we can give recommendation to the future customers and it will be very helpful to them for buying the products [2].

Designing Ranking Systems for Consumer Reviews the Impact of Review Subjectivity on Product Sales and Review Quality provide two ranking systems: a consumer-oriented ranking mechanism ranks the reviews according to their expected helpfulness, and a manufacturer-oriented ranking mechanism ranks the reviews according to their expected effect on sales [5]. From all our reference the customers were not satisfied because of ranking and comparison of different E-commerce sites.

The customer compares different products based on prices and schemes available on that particular product given by different E-commerce websites. As soon as the customer clicks on the buy button on the product suitable according to him, this triggers the customer to original E-commerce website from where customer can do original purchasing of the product [7].

### III. SYSTEM ARCHITECTURE.

Fetch the reviews from different website for a single product there may be more than 1000 reviews from the same product. An Internet Bot, also known as web robot, WWW robot or simply bot, is a software application that runs automated tasks (scripts) over the Internet. Typically, bots perform tasks that are both simple and structurally repetitive, at a much higher rate than would be possible for a human alone. The largest use of bots is in web spidering, in which an automated script fetches, analyzes and files information from web servers at many times the speed of a human, by the help of dictionary we collect all positive and negative comments from users review. All comments have its own score accordingly it displays as positive and negative. From the fetching reviews it stores all the

details about the product example: product mobile (Mobile company, Mobile Model Number, Screen Resolution, Camera Quality, Battery Power, Price and so on) in the database with its review {Good, Bad}. The database gets more number of tuples from different E-commerce sites, and it is possible to get data Redundancy. So we have to use data mining technique to remove the redundancy. By using Data Mining algorithms, it gets started to analyze the database and display the supervised datasets. Finally, it monitors whether the product are Good or Bad, based on consumer budget by the help of decision making system.

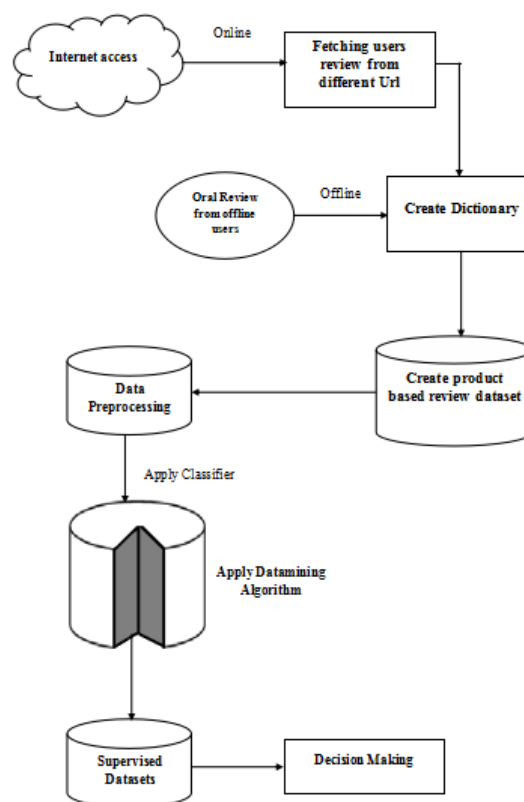


Fig 1: System Architecture

The Product Analysis using Data Mining and it solves the various problems of user by providing the user Products from different E-commerce url at one place with different prices and schemes and offers that are offered by different E-commerce firms. This will provide the user privilege to choose products from different E-commerce applications which they consider is the best for them and their reviews. For this, we make use of machine Learning algorithms that keep Track of the user's behavior and searching patterns.

#### IV. WORKING PRINCIPLES

The quality of data get analyzed by system, from different E-commerce sites with user's opinion. Huge amount of data can fetch from each url for a single product. The common attributes will manage all different features from different users online and offline opinion. Here various steps which helps to moderate the system for Monitoring the Best product from online and offline user's opinion and paralely we analyse the product price which the users get satisfy from this system. The different steps where mentioned from figure (1).

##### A. Internet Bot

An Internet bot, in its most generic sense, is software that performs an automated task over the Internet. More specifically, a bot is an automated application used to perform simple and repetitive tasks that would be time-consuming, mundane or impossible for a human to perform. One of the best examples of a good bot is a search engine spider. Such bots troll the Web and index new pages for a search engine. A gray area also exists where bots are used by websites to promote shopping deals. In such cases, a bot will perform much like a search engine spider to index information about products on the Web. Some e-commerce sites take steps to ban shopping bots of this type, while their creators feel that they are simply providing a better service to their own users. The largest use of bots is in web spidering (*web crawler*), in which an automated script fetches, analyzes and files information from web servers at many times the speed of a human. More than half of all web traffic is made up of bots.

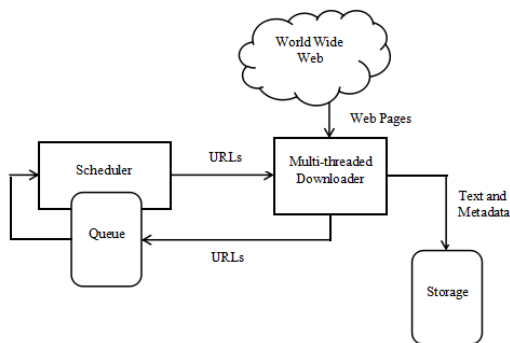


Fig 2: Internet Bot

##### B. Create Product based datasets:

The large amount of data collected from different web sites and normal offline reviews, by using this reviews we create the dictionary for developers knowledge to ensure that what type of reviews it. Each review has its score if the score is high for a review then we assume the review is positive, if the review score is low then we assume the review is

bad. Review is the class attribute for product dataset. Customers can review a product based on its specification, they give subjective reviews about the product in which some will positive and some will negative, the system will point out according to its score. It stores both numeric and nominal data, the last attribute is class and we categories it as good, bad. Orally it scores the review from offline customers. Example: Product Mobile Sony Xperia XAPros

1. Premium Design, attractive look !!
2. 13 MP main camera with LED flash!! click amazing pictures !!!
3. 8 MP front camera!! It clicks nice selfies !!
4. The internal storage is 16 GB and the internal storage iz expandable up to 200 GB !! Can load many apps and games!!!
5. Android 6.0.1 Marshmallow (latest android version)
6. Predictive Hybrid Auto focus cameras!!
7. Supports 4G!!
8. 2.5D curved glass. Very good protection to the display.

The above underline words are positive reviews, we collect these words and score them as 5 out of 5, then we assume that the score above 3 as good and below 3 as bad. More number positive as well as negative reviews will pic from different E-commerce sites. The above review from given url [http://www.amazon.in/Sony-Xperia-XA-Graphite-Black/product-reviews/B01H4U80R8/ref=cm\\_cr\\_ar\\_p\\_d\\_viewpnt\\_lf\\_t?showViewpoints=1&filterByStar=positive&pageNumber=1](http://www.amazon.in/Sony-Xperia-XA-Graphite-Black/product-reviews/B01H4U80R8/ref=cm_cr_ar_p_d_viewpnt_lf_t?showViewpoints=1&filterByStar=positive&pageNumber=1)

##### C. Data Preprocess:

Data preprocessing is a data mining technique that involves transforming raw data into an understandable format. Real-world data is often incomplete, inconsistent, and/or lacking in certain behaviors or trends, and is likely to contain many errors. Data preprocessing is a proven method of resolving such issues. From the datasets there may be the possibilities of same tuples in multiple time so it may confuse to move further. So we choose data preprocessing to make redundancy. It will help to reduce the datasets with removing duplicate tuples.

##### D. Datamining Algorithm:

An algorithm in data mining (or machine learning) is a set of heuristics and calculations that creates a model from data. To create a model, the algorithm first analyzes the data you provide, looking for specific types of patterns or trends. The algorithm uses the results of this analysis over many iterations to find the optimal parameters for creating the

mining model. These parameters are then applied across the entire data set to extract actionable patterns and detailed statistics.

Choosing the best algorithm to use for a specific analytical task can be a challenge. While we can use different algorithms to perform the same business task, each algorithm produces a different result, and some algorithms can produce more than one type of result. For example, using Decision Trees algorithm not only for prediction, but also as a way to reduce the number of columns in a dataset, because the decision tree can identify columns that do not affect the final mining model.

However, there is no reason to choose limited algorithm for the solutions. Experienced analysts will sometimes use one algorithm to determine the most effective inputs (that is, variables), and then apply a different algorithm to predict a specific outcome based on that data. Monitoring the best product from consumer's opinion is to decide whether to buy a product, datamining algorithm will help to give more satisfaction to the users. In our proposed system we choose data classifier as an algorithm to monitor the best product.

#### E. DataClassifier:

A classifier is a Supervised function (machine learning tool) where the learned (target) attribute is categorical ("nominal"). It is used after the learning process to classify new records (data) by giving them the best target attribute.

There are several challenges in working with data classification. One in particular is that it is necessary for all using categories on e.g. customers or clients, to do the modeling in an iterative process. This is to make sure that change in the characteristics of customer groups does not go unnoticed, making the existing categories outdated and obsolete, without anyone noticing.

The first step in doing a data classification is to cluster the data set used for category training, to create the wanted number of categories. An algorithm, called the *classifier*, is then used on the categories, creating a descriptive model for each. These models can then be used to categorize new items in the created classification system. In our system we choose Naive Bayes Algorithm for analysing the best product from customer reviews. In machine learning, naive Bayes classifiers are a family of simple probabilistic classifiers based on applying Bayes' theorem with strong (naive) independence assumptions between the features.

Naive Bayes is a simple technique for constructing classifiers: models that assign class labels to problem instances, represented as vectors of feature values, where the class labels are drawn from some finite set. It is not a single algorithm for training such classifiers, but a family of algorithms

based on a common principle: all naive Bayes classifiers assume that the value of a particular feature is independent of the value of any other feature, given the class variable. For some types of probability models, naive Bayes classifiers can be trained very efficiently in a supervised learning step. In many practical applications, parameter estimation for naive Bayes models uses the method of maximum likelihood; in other words, one can work with the naive Bayes model without accepting Bayesian probability or using any Bayesian methods. Naive Bayes algorithm is support for analysing the best product consider the product dataset as a training dataset and apply naive bayes algorithm for the particular tuple.

NB makes predictions using Bayes' Theorem, which derives the probability of a prediction from the underlying evidence. Bayes' Theorem states that the probability of event A occurring given that event B has occurred ( $P(A|B)$ ) is proportional to the probability of event B occurring given that event A has occurred multiplied by the probability of event A occurring ( $(P(B|A)P(A))$ ).

#### F. Supervised Datasets:

Classification of a collection consists of dividing the items that make up the collection into categories or classes. In the context of data mining, classification is done using a model that is built on historical data. The goal of predictive classification is to accurately predict the target class for each record in new data, that is, data that is not in the historical data. In our system we choose the common model for same product with its all specification, from preprocessed dataset, in which the different reviews may occur in class attribute from different E-commerce site. To select same products with different review's from preprocessed datasets by using sql queries. This has to follow for all same product and its models then each model of the product we consider as a file or dataset. Naive Bayes algorithm is to find the probability of each dataset, Example {Sony, m1} this reviews from different websites and different customers offline reviews.

S.no	Brand	model	Screen Resolution	Camera Resolution	Audio/ Video	Battery Capacity (mAh)	price	rating	review
1	Sony	m1	High	High	Yes	3500	10000	4	Good
2	Sony	m1	High	High	Yes	3500	13000	4	Avg
3	Sony	m1	High	High	Yes	3500	9000	3	Avg
4	Sony	m1	High	High	Yes	3500	13000	4	Avg
5	Sony	m1	High	High	Yes	3500	10000	4	Good
6	Sony	m1	High	High	Yes	3500	9000	3	Avg
7	Sony	m1	High	High	Yes	3500	9000	3	Avg
8	Sony	m1	High	High	Yes	3500	13000	4	Avg
9	Sony	m1	High	High	Yes	3500	10000	4	Good

Fig 3: Supervised Dataset for Single product

The above example table product, model and its specification are same but price, rating and reviews are different, because each different tuple are different opinion from customers. So we apply Naive Bayes algorithm to find the highest probability from class attribute to further decision. Using weka the class 'Avg' is the highest probability which has the average 0.64. From this the least price tuple will recorded in final database like the same for all different model from same product.

### **G. Decision Making:**

Information got is what is transmuted to decisions. Decisions are taken by decision takers who have certain organizational objectives in mind with a certain way of processing and appreciating information. That is why the right information has to be supplied to decision takers in an organization.

Some people take in information best if it highly detailed and specific. The various elements of information need not to be linked as a whole. The other group absorbs information in a holistic way, that is in a less concrete way, preferring general facts, and softdata linked as a whole.

They further postulated that the group that absorbs the information in totality will be involved in a high degree of analytic thought in providing adequate and detailed justifications involving quantitative reasons in support of final decisions. While the other group will rely more on intuition, experience, rules of thumb and judgement making it difficult for this group to provide justification for recommended decisions.

Here this system will monitor the best product through datamining algorithm and give satisfied reviews based on customers budget. Finally, all the data has been recorded in new database, user can trust the reviews and they can decide to purchase the product. From users point of view they gives product name and budget range, system will gives details with trusted reviews about the product based on their budget.

### **V. FUTURE ENHANCEMENT**

In the future, we aims to include several features like suggesting the users feedback about this review and their rating about the expected system. It display the every years report with product updated version. Statistical analysis for each product should be include in it. It will increase the business stratagy and demand of the product in market.

### **VI. CONCLUSION**

The rapid growth of the Internet, users' ability to post the review after purchasing any product. They

post different types of comments and there Opinion in the review column, we contribute the existing research that it has only the review with different e-commerce sits and To obtain the best deals from different E-commerce websites, a web crawler is used to crawl on different E-commerce websites and fetch the URLs of products. The scrapper scrapes the details abstracted within the URLs and stores it in the database. Then comparison among products of different E-commerce websites is made by using techniques such as inverted indexing. Here we monitor the best product to purchase from online or offline based on consumers budget through E-commerce sits or direct purchase. From this system user can get clear idea about the product and they can decide to purchase or not.

### **REFERENCES**

- [1]. Jawahire Nakash. 'Real Time Product Analysis using Data Mining', International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), Volume 4 Issue 3, March 2015.
- [2]. Sushant Bhagwat . 'Sales Analysis using Product Rating in data mining Techniques', IJRET: International Journal of Research in Engineering and Technology , Volume: 04 Issue: 02 | Feb-2015 eISSN: 2319-1163 | pISSN: 2321-7308.
- [3]. Seddin K, Serenko and Hayes J , 'Online shopping bots for E-commerce', Int. J. Electronic Business, pp. 556-589, (2007)
- [4]. Chevalier, J., and D. Mayzlin. 2006. The effect of word of mouth online: Online book reviews. Journal of Marketing Research, forthcoming.
- [5]. Anindya Ghose . ' Designing Ranking Systems for Consumer Reviews: The Impact of Review Subjectivity on Product Sales and Review Quality ', Department of Information, Operations, and Management Sciences Stern School of Business New York University.
- [6]. Clark, D. 'Shopbots become agents for business change', IEEE Computer, Vol. 33, Issue 2, 2000.
- [7]. Serenko and Hayes J 'Investigating the functionality and performance of online shopping bots for E-commerce', Electronic Business 8(1): pp.1-15,(2009).