IJCSE International Journal of Computer Sciences and Engineering Open Access

Review Paper Vol.-6, Issue-8, Aug 2018 E-ISSN: 2347-2693

# Opinion Mining from Customer Reviews for Product Ranking

**Jahiruddin**

Dep. of Computer Science, Jamia Millia Islamia, New Delhi, India

*Corresponding Author: jahir.jmi@gmail.com, Tel.: +91-98715-07112*

**Available online at: www.ijcseonline.org**

Accepted: 17/Aug/2018, Published: 31/Aug/2018

***Abstract***— Recently the peoples of the metropolitan cities are moving from traditional offline interactive shopping to online shopping due to time limitation and cost of products. In online shopping, the purchase decision is a challenging task for new customers as there may a large number of competitive products. Recently mostly online shopping sites have been facilitated to their customers to write the reviews about the products they have purchased. These customers' reviews do not only help to new customer for taking purchase decision but also help the manufacturer to increase the sale of their products by improving its quality. This paper presents a reviews mining method to extract product features and its opinion. Thereafter, we apply the Analytic Hierarchy Process (AHP) on extracted features and opinion to rank the competitive products by scoring them. The method has been validated on a data set related to five smart phones downloaded from three deferent online shopping websites - *Flipkart, Snapdeal,* and *Amazon*. The evaluation result shows that the proposed method gives up to marks result.

***Keywords***— Text mining; Opinion mining; feature extraction; Analytical Hierarchy Process; Product ranking

## I. Introduction

Due to easy availability of online shopping websites and time problems in metropolitan cities, the customers are shifting from traditional interactive shopping to online shopping to purchase products at completive cost. The overall cost of product in online shopping is low in comparison to interactive shopping as the product generally sells directly to customers from manufacturer and customer are able to access product price and details from different online shopping websites and compare them easily to take purchase decision. The customer feel very difficult to take purchase decision based on product descriptions and prices on different online websites but they needs feedback of the customers who had already purchased it at the time of online product purchasing.

Recently, majority of the online shopping website allows their customers to write reviews about the products they had purchased. The opinion of the existing customers are reliable and important documents that help new customers in taking purchase decision and it also help to manufacturer to improve quality of their products to increase overall sale of the product. But due to distribution of review documents of same product across a number of online websites and as it is in textual form, it is a challenging task to mine the customers' reviews and quantify it to rank the products using these documents that may be used by new customer. Though few exiting websites like *mysmartprice.com*, *maaptol.com* etc. compare the similar products based on their main features and price, but to the best of my knowledge none of the website uses the customers' reviews in products comparison and ranking.

In this paper, we proposed a method for opinion mining that extract the features and opinion word of customers about a product from review documents. Only features and opinions are not helpful to new customer in taking purchase decision, therefore after opinion mining we have identified five features – *price, camera, battery, screen,* and *phone* to rank the products of similar kind by calculating rank score of each product using Analytic Hierarchy Process (AHP). The AHP is a Multi Criteria Decision Making (MCDM) technique which is used to rank the features and products.

The rest of the paper is organization as follows. Section 2 presents a brief review of the existing works on opinion mining and product ranking using different approaches. The functioning details of our proposed method are presented in section 3. Section 4 presents the experimental and evaluation results of our proposed method. Finally, section 5 concludes the paper.

## II. Related Works

The customers' reviews are very important and reliable documents, which may help both the customers in taking the purchase decision and manufacturer to improve the sale of their products by improving its quality for customers' satisfaction. Since our proposed work is based on opinion mining followed by product ranking, in this section, we present a brief works on opinion mining from customers' reviews followed by products ranking.

A large number of researchers working in opinion mining have been targeted to extract features of products along with opinion bearing words e.g., *great, awesome, good, poor, bad, very poor* etc. In literature, a good number of researchers have been attempted to mine such words along with their semantic meaning [1, 2]. Another work in the area of opinion mining is *"sentiment analysis"*, which attracts a reasonable number of researchers [3, 4]. An acceptable number of papers on *sentiment analysis* focus on the customers' reviews classification such as – positive or negative [5, 6], subjective or objective [7] etc. Although, the *sentiment analysis* works are able to classify the review documents into positive, negative, neutral, or subjective, objective classes but it fails to find reviewers overall opinion about the products. A negative document on a product does not show that the reviewer dislikes the product with respect to all features or aspects of the product; similarly a positive document for same product does not mean that the reviewer has positive opinion with respect to every feature of the product. Generally, the customer writes both positive and negative points of the product along with overall sentiment on the product. To get the overall sentiment about a product along with detailed aspects, feature-opinion mining is proposed in a number of literatures [8, 9, 10, 11]. In [8], the authors proposed a framework to compare the products by analysing customers' reviews with respect to a number of identified features of the competing products. But it does not address the ranking of product by scoring them. In [9, 10], the authors proposed lexicon based opinion mining approach for customers' reviews. They reported that the performance of their approach in opinion mining is quite well. In [11], the authors presented an unsupervised customers' review mining system which extract the product features and its opinions. In [12], the authors presented various opinion mining problems along with a number of opining mining techniques. They also address to detect the fake products reviews and opinion spam. The sentiment analysis and product reviews summarization is also attracted a number of researchers [13, 14, 15].

In order to identify customer's preferences about features of a smart phone, some researchers have worked in this direction [16, 17, 18, 19, 20, 21]. In [17], the authors presented that the customer's requirements play a vital role in the success of a new product. The product ranking attracted a number of researchers [22, 23, 24, 25], but to the best of our knowledge, no worked have been done on product ranking using feature and opining mining from customers' reviews.

## III. PROPOSED OPINION MINING AND PRODUCT RANKING METHOD

In this section we describe the different modules of our proposed opinion mining and product ranking method. The aim of the proposed method is to extract features and opinion from customers' review documents that may be used in product ranking by generating the rank score for each product in a given set of competitive products of same king. Figure 1 shows the functional details of our proposed method. It starts by creating a data set of customers' reviews at local machine using data *crawling and pre-processing* module. Thereafter, it identifies and ranks the important and common features of the products under considerations in *Products' Feature Identification and Ranking* module. The *Opinion Mining* module is used to extract feature and opinion for each product using their customers' review documents. Finally, the products are ranked by generation their rank score in *Product Ranking* module. Following sub-section explain the functional details of these modules.

### *A. Data Crawling and Pre-processing*

In order to rank the online product by opinion mining using review documents, it is needed to create a data set of customers' reviews on local machine. We have downloaded review documents from three popular online shopping websites *Flipkart*, *Snapdeal*, and *Amazon*. The review document have a number of product's information like *price, star rating, review-title, review-content,* etc along with customer's information such as – *user name, post date, user verification status,* etc. But in our opinion mining and product ranking method we considered only *review-titles* and *review-content*. Table 1 shows sample customer reviews on *Samsung Galaxy S7* and *HTC Desire 10 Pro* smart phone downloaded from three online shopping websites mentioned above.

### *B. Product's Feature Identification and Ranking*

In this module we have to identify important and common features of the product from a large list of product's features. In our cases of the smart phones, we have identified five features – *price, camera, battery, screen*, and *Phone*. After feature identification, it is needed to rank them by calculating the rank score for each feature. We rank the features using Analytic Hierarchy Process (AHP) [26]. The feature ranking process using AHP has following steps:

*Step 1: Features' Relative Score Matrix Generation*

In order to rank the features, the first step is to generate the features' relative score matrix $F$. For $n$ number of features, the $F$ should be an $n \times n$ reciprocal matrix. The elements of this matrix are obtained by domain expert using the Saaty's nine-point scale given in table 2, where $f_{i,j}$ and $f_{j,i}$ must be reciprocal to each other.

*Step 2: Feature Ranking and Consistency Checking*

After getting the matrix *F*, we will rank the features by calculating the principle eigenvector of the matrix *F*. There are a number of methods to calculate the principle eigenvector of a matrix, we calculate it first dividing each column elements by column-sum then taking the average of each row as mentioned in [27]. Let this principle eigenvector is represented by an *n* dimensional vector *S*, where $s_i$ is the rank score of feature $f_i$. Besides, feature ranking AHP also provide method to check whether the matrix *F* is consistent or not. For consistency checking, first we calculate $A = F \times S$ which results an *n* dimensional vector *A*. Next we divide each element of *A* by corresponding element of *S* and get the value of λ by taking the average of this resultant vector. Next we calculate the *CI (consistency index)* using equation 1 and then compute the ration *r* between *CI* and *RCI (Random Consistency Index)* using equation 2. A partial list of value of *RCI* is presented in table 3 given by the author in [27]. The matrix *F* is consistent, if $r < 0.1$, otherwise it is inconsistent, and new matrix *F* is generated using expert's new feature's relative scores.

$$CI = \frac{(\lambda - n)}{n-1} \tag{1}$$

$$r = \frac{CI}{RCI} \tag{2}$$

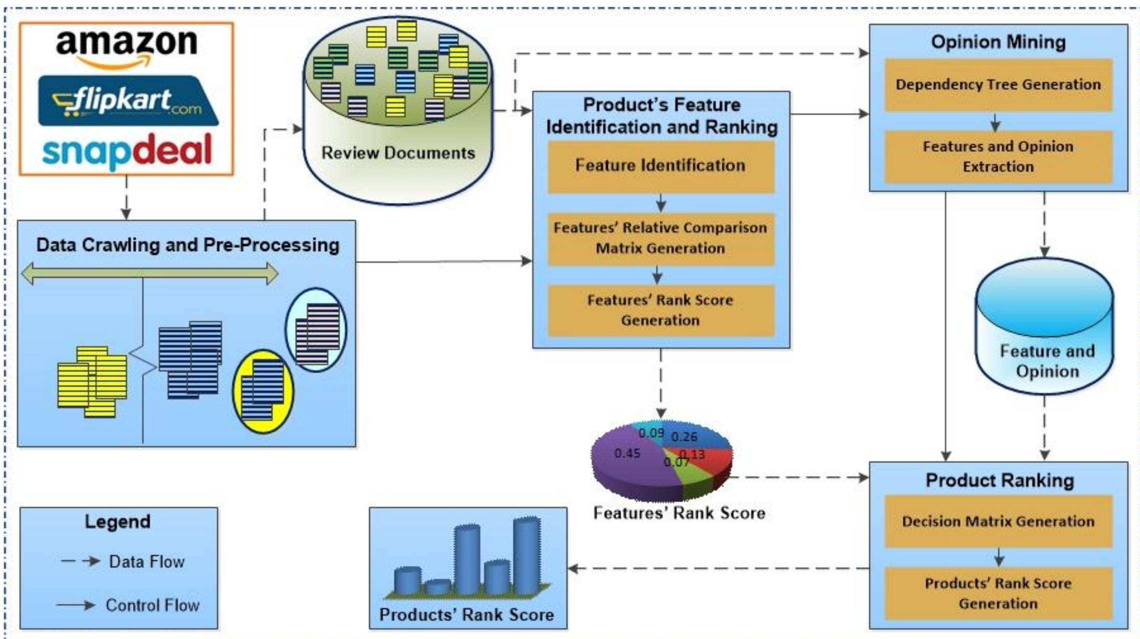


**Figure 1:** Functioning details of the proposed method

**Table 1:** Sample reviews related to Samsung Galaxy S7 and HTC Desire 10 Pro smart phone

| Smart Phones | Source | Review Title | Review Content |
|---|---|---|---|
| Samsung Galaxy S7 | Amazon | price is too high Don't buy | Price is too high pls wait for sometime soon the price will be lower down Or u can go for s8 |
| Samsung Galaxy S7 | Snapdeal | Samsung Galaxy S7 Edge | Best phone of 2016 as many youtubers made it as a daily driver. |
| Samsung Galaxy S7 | Flipkart | Great phone | Value for money !; great battery; great display; and great camera ! |
| HTC Desire 10 Pro | Amazon | camera is not good | To be very honest selfie camera sucks.; even a 5 mp camera is better and the rear camera is also not that good; look wise phone is nice |
| HTC Desire 10 Pro | Snapdeal | Excellent Product | good phone with fair battery life |
| HTC Desire 10 Pro | Flipkart | Worth the money | Great camera but little expensive |

**Table 2:** Saaty's nine point scale for feature-pair $f_1$ and $f_2$

| Linguistic text | Score |
|---|---|
| $f_1$ and $f_2$ are equally important features | 1 |
| $f_1$ is slightly more important feature than $f_2$ | 3 |
| $f_1$ is more important feature than $f_2$ | 5 |
| $f_1$ is strongly more important feature than $f_2$ | 7 |
| $f_1$ is extremely more important feature than $f_2$ | 9 |
| Relative importance in between above mentioned | 2, 4, 6, 8 |

**Table 3:** List of *Random consistency Index (RCI)* for $n = 2$ to $n = 8$

| n | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|
| RCI | 0.00 | 0.58 | 0.90 | 1.12 | 1.24 | 1.32 | 1.41 |

### C. *Opinion Mining*

In order to extract the features and opinions of the products from review documents, we have written a rule based program in Java. To get the rules for <*feature, modifier, opinion*> triplet extraction we have used following steps:

*Step 1:* We manually identified the sentences that have the features and opinions.

*Step 2:* We extracted <*feature, modifier, opinion*> triplets from these sentences by manually analysing them.

*Step 3:* Dependency tree for each sentence is generated, and then frame the rules to get the extracted triplets in step 2.

*Step 4:* Rules are implemented and run on whole corpus. For a given rule, if it results more *true positive* triplets than *false positive* then it should be retains otherwise dropped.

The figure 2 shows the list of identified rules for <*feature, modifier, opinion*> triplets extraction. The <*feature, modifier, opinion*> triplets extraction program takes dependency tree and corresponding *word and tags* of a sentence and return the triplets if exist in it. We have used Stanford parse[1] to generate dependency tree and POS tags to each word of a sentence. The rule 2 and its sub-rules may results a large number of false positive triplets, so to overcome this problem we have filtered out a triplet if its *feature* component does not consists a target feature.

Figure 3 presents the dependency tree along with POS tags and extracted triplet <*price, too, good*> from the sentence *"price is too high Don't buy."* using rule 1.3. In this figure the participated nodes are coloured with green colour and corresponding links with thick line. The Figure 4 shows that from the sentence *"great battery; great display; and great camera."* The triplets <*batter, null, great*> and <*camera, null, great*> are extracted using rule 2, but the triplet <*display, null, great*> is missing due to miss-tagging *display* as *VBP* by the parser. Our program is also able to get the negative modifier of an opinion see figure 5. Figure 6 shows that triplets <*phone, null, good*> and <*battery life, null, fair*> are extracted by applying *rule 2* and *rule 2.2* respectively on sentence *"good phone with fair battery life."*

### D. *Product Ranking*

In order to rank the competitive products using AHP, first we have to generate the decision matrix. The decision matrix $D$ is an $m \times n$ real matrix, where $m$ is the number of competitive products and $n$ is the number of target features. In decision matrix generation process it takes valid extracted <*feature, modifier, opinion*> triplets for a given feature of the product and get a real number by taking the average of the corresponding *modifier + opinion* words' numeric sentiment score. A triplet is valid if its feature part contains a target feature and opinion along with modifier should be a possible opinionated word for corresponding feature.

After generation of the decision matrix $D$ we calculate the $R = D \times S$ which results an $m$ dimensional vector $R$. The *ith* element of the vector $R$ is the rank score of the *ith* product. The rank score of a product is simple the sum of the products of different feature value of the product in matrix $D$ and corresponding feature rank score of the vector $S$. This ranking process not only ranks the product it also calculate rank score of each product.

## IV. EXPERIMENTAL SETUP AND RESULTS

For experimental evaluation of our proposed opinion mining and product ranking method, we created a dataset of 5623 review documents which are downloaded from three popular online shopping websites such as *Flipkart, Snapdeal*, and *Amazon* related to five competitive smart phones *iPhone 7, Samsung Galaxy S7 Edge, Google Pixel, HTC Desire 10 Pro,* and *Lenovo Z2 Plus*. The summary of this data set is presented in table 4.

[1] http://nlp.stanford.edu/software/lex-parser.shtml

 

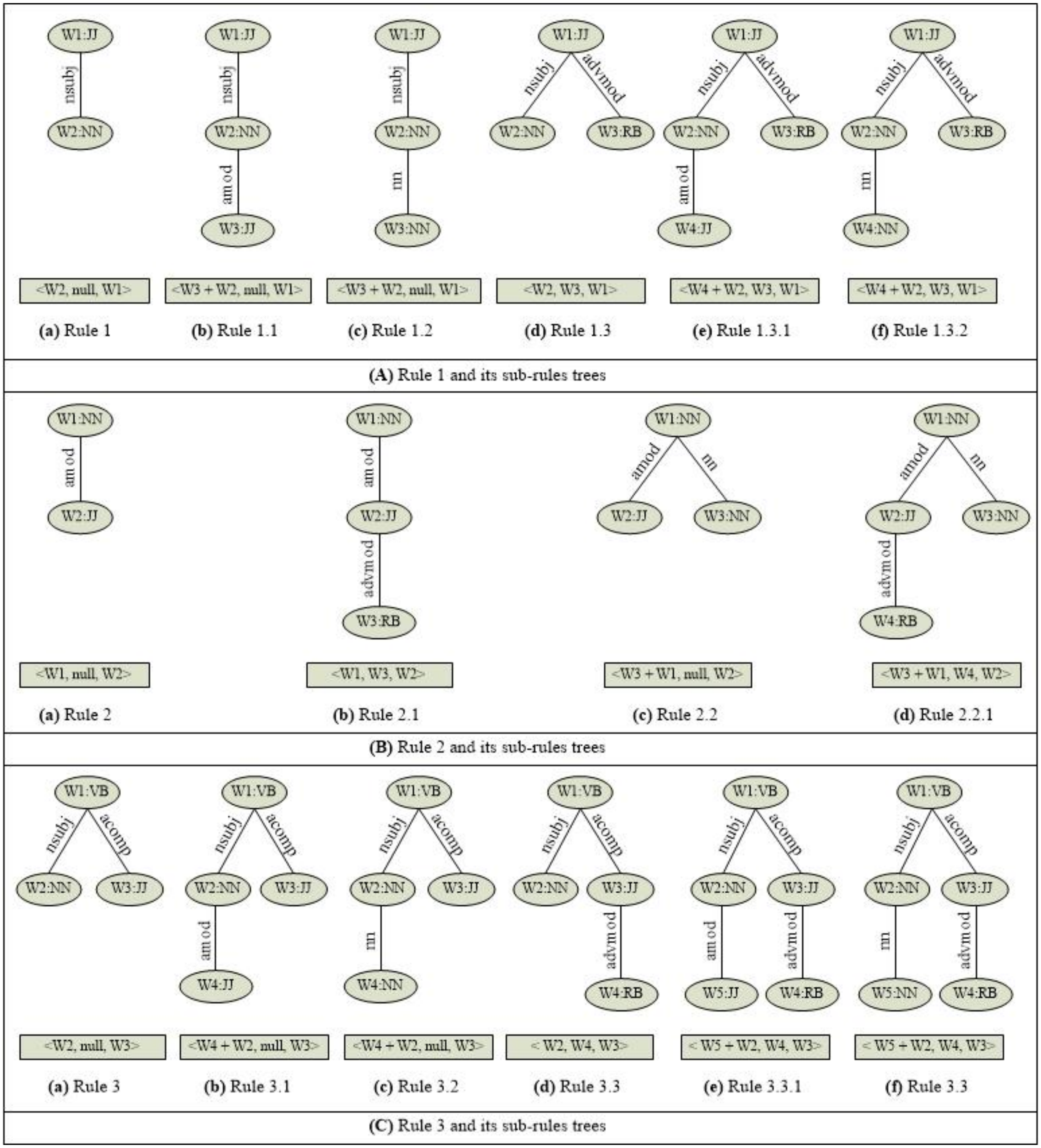


**Figure 2:** Set of rules to extract <feature, modifier, opinion> triplets

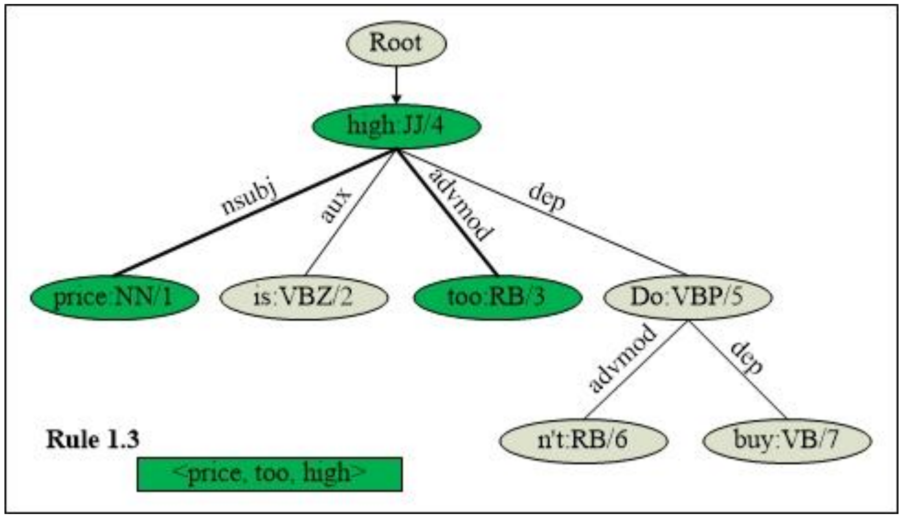


**Figure 3:** dependency tree and extracted triplet from sentence *"price is too high Don't buy."*

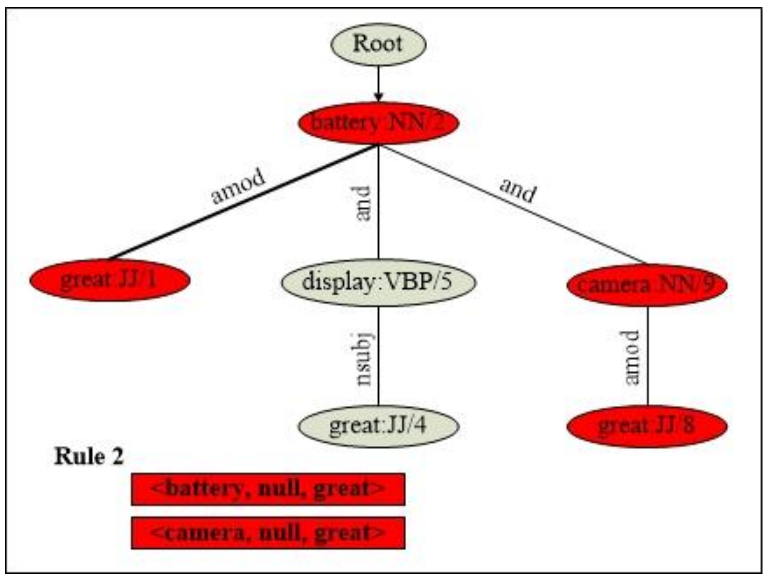


**Figure 4:** dependency tree and extracted triplet from sentence *"great battery; great display; and great camera."*

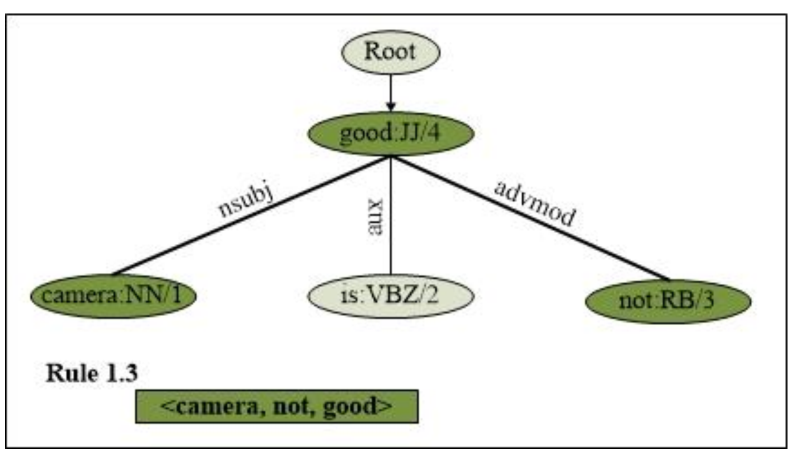


**Figure 5:** dependency tree and extracted triplet from sentence *"camera is not good."*

 

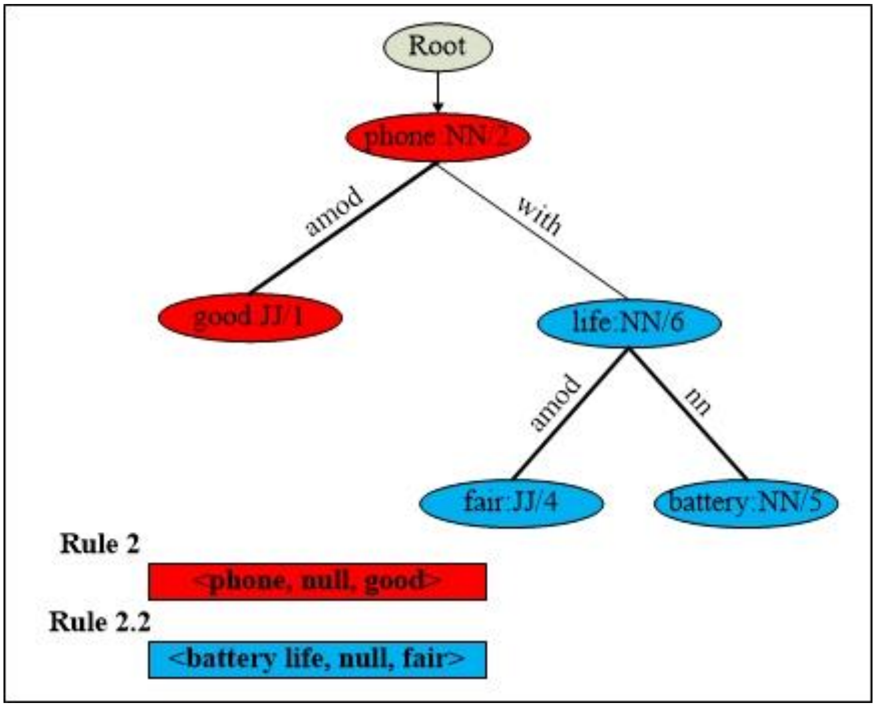


**Figure 6:** dependency tree and extracted triplet from sentence *"good phone with fair battery life."*

**Table 4:** Summary of the data set on smart phones

| Smart Phone | Number of Review documents downloaded from | | | Total Review Documents |
|---|---|---|---|---|
| | Flipkart | Snapdeal | Amazon | |
| iPhone 7 | 1116 | 121 | 702 | 1939 |
| Samsung Galaxy S7 Edge | 135 | 3 | 307 | 445 |
| Google Pixel | 180 | 9 | 113 | 302 |
| HTC Desire 10 Pro | 120 | 4 | 100 | 224 |
| Lenovo Z2 Plus | 310 | 224 | 2179 | 2713 |

After creation of the data set on local machine, the next step in this method is feature identification and ranking. With the help of domain experts we have identified five features – *price, camera, battery, screen,* and *Phone* which are common for all competitive products. Next we have generated Features' relative comparison matrix by assigning relative score of each feature-pair. Since we have *five* targeted features so total *(5 x 4)/2 = 10* relative score is needed. Table 5 present these relative score for each feature-pair and corresponding features' score matrix is shown in figure 7. Next we normalize the matrix *F* by dividing each element of a column by its column sum and take the average of each row of this normalized matrix to get the feature score vector *S*. The figure 8 shows the normalized matrix *F'* and feature score vector *S*. From vector *S* we gets that the feature *phone* is rank first with rank score *0.5377* and *price* ranked second with rank score *0.2613*. The rank and rank score of identified five features are shown in table 6.

**Table 5:** Relative Score of each Feature-pair of target features

| Relative preferences of feature-pair | Score |
|---|---|
| *Price* is strongly more important than *camera* | 7 |
| *Price* is more important than *battery* | 5 |
| *Price* is more important than *screen* | 5 |
| *Price* is less important than *Phone* | 1/5 |
| *Camera* is slightly less important than *battery* | 1/3 |
| *Camera* is slightly less important than *screen* | 1/3 |
| *Camera* is strongly less important than *Phone* | 1/7 |
| *Battery* and *screen* are equally important | 1 |
| *Battery* is strongly less important than *Phone* | 1/7 |
| *Screen* is strongly less important than *Phone* | 1/7 |

 

$$F = \begin{bmatrix} 1 & 7 & 5 & 5 & 1/5 \\ 1/7 & 1 & 1/3 & 1/3 & 1/7 \\ 1/5 & 3 & 1 & 1 & 1/7 \\ 1/5 & 3 & 1 & 1 & 1/7 \\ 5 & 7 & 7 & 7 & 1 \end{bmatrix}$$

**Figure 7:** Features' Score matrix *F*

$$F^{'} = \begin{bmatrix} 0.1528 & 0.3333 & 0.3488 & 0.3488 & 0.1228 \\ 0.0218 & 0.0476 & 0.0233 & 0.0233 & 0.0877 \\ 0.0306 & 0.1429 & 0.0698 & 0.0698 & 0.0877 \\ 0.0306 & 0.1429 & 0.0698 & 0.0698 & 0.0877 \\ 0.7642 & 0.3333 & 0.4884 & 0.4884 & 0.6140 \end{bmatrix} \quad S = \begin{bmatrix} 0.2613 \\ 0.0407 \\ 0.0801 \\ 0.0801 \\ 0.5377 \end{bmatrix}$$

**Figure 8:** Normalized matrix *F'* and feature score vector *S*

**Table 6:** Feature and their rank score calculated using AHP

| Feature | Rank | Rank Score |
|---|---|---|
| Price | 2 | 0.2613 |
| Camera | 4 | 0.0407 |
| Battery | 3 | 0.0801 |
| Screen | 3 | 0.0801 |
| Phone | 1 | 0.5377 |

In order to check consistency of matrix *F*, first we compute the $A = F \times S$ then we divide each element of vector *A* by corresponding element of vector *S* that gives a 5 dimensional vector *B*. The value of λ which is used in calculation of *CI* is obtained by taking the average of vector *B*. Finally we calculated the ration $r = CI/RCI$. Figure 9 shows the calculation of $CI / RCI$ ration *r* where $RCR (5) = 1.12$ (see table 3). The value of $r = 0.0893$ which is less than *0.1* therefore the matrix *F* is consistent and features' scores are acceptable.

After feature ranking using AHP which calculates the rank score of each identified feature, we have been ranked the competitive products using AHP. In product ranking first of all we have generated the decision matrix. In decision matrix generation, we have used valid <*feature, modifier, opinion*> triplets for each target *feature* related to each product. Since the decision matrix is a real matrix and *opinion* along with *modifier* is text, so it is needed to convert each opinion word into a real value. We have used the *textBlob* python module in this conversion process. The *textBlob* is a natural language tool written in python that may be used to get the sentiment polarity of a text. In some cases it is unable to assign correct sentiment polarity score, so experts help is needed to correct such score. For example in our case, the *sentiment polarity* scores, assigned by *textBlob*, of the opinion words *"too high"* and *"very high"* for the feature *price* are 0.16 and 0.21 respectively but it should be -1.0. Table 7 shows the possible opinion words along with their *sentiment polarity score* corresponding to each feature. To get the decision matrix we have converted opinion word of each extracted triplet corresponding to each feature of a product and take average of this score to get the entry of decision matrix. The table 8 shows the decision matrix *D* of our data set. Since there are five *products* and five *features*, therefore the order of the decision matrix *D* in our case is *5 x 5*.

After Decision matrix generation, we multiply it by feature score vector *S* to get the smart phones' rank score vector *R*. Figure 10 shows the steps for calculation of *R*. From vector *R* we gets that the *iPhone 7* is rank first with rank score *2.07* followed by *Samsung Galaxy S7 Edge* ranked second with rank score *1.92*. Table 9 shows the ranks of the smart phones with their rank score and visualization of it shown in figure 11.

### A. Evaluation Result of Ranking Process

In this section, we present the evaluation result of the ranking process of our proposed *opinion mining and product ranking method*. Since there are no benchmark data sets of smart phones that have relative ranks of different smart phones, we have taken rank of these competitive smart phones by two domain experts and calculating the overlapping score using *set intersection* method [28]. Table 10 shows the rank lists $L_1$ and $L_2$ of different smart phones ranked by two domain experts and *L* ranked by our proposed method. Table 11 present the calculation of overlapping scores of *L* with $L_1$ and $L_2$ using set intersection method. From this table we get that aggregate average overlapping score of *L* with $L_1$ and $L_2$ is 85.0%, which show that products' ranking result of our proposed method is closer to the experts' rank. Therefore, it may be used to calculate the ranks of the competitive products from customers' reviews that may be help both for new customers in purchase decision and manufacturer to improve quality of their products.

 

$$A = F \times S = \begin{bmatrix} 1 & 7 & 5 & 5 & 1/5 \\ 1/7 & 1 & 1/3 & 1/3 & 1/7 \\ 1/5 & 3 & 1 & 1 & 1/7 \\ 1/5 & 3 & 1 & 1 & 1/7 \\ 5 & 7 & 7 & 7 & 1 \end{bmatrix} \times \begin{bmatrix} 0.2613 \\ 0.0407 \\ 0.0801 \\ 0.0801 \\ 0.5377 \end{bmatrix} = \begin{bmatrix} 1.46 \\ 0.21 \\ 0.41 \\ 0.41 \\ 3.25 \end{bmatrix}; \quad B = \begin{bmatrix} 5.57 \\ 5.11 \\ 5.14 \\ 5.14 \\ 6.05 \end{bmatrix};$$

$$\lambda = 5.40; \quad CI = \frac{(5.40-5)}{(5-1)} = 0.10; \quad r = \frac{0.10}{1.12} = 0.0893$$

**Figure 9:** Steps for calculation of ration $r = CI / RCI$

**Table 7:** List of Possible Opinion Words with their Sentiment Polarity Score for each Feature

| Feature | List of Possible Opinion Words along with sentiment polarity score |
|---|---|
| Price | Amazing:1.0, Awesome:1.0, Bad:-0.69, Best:1.0, Cheap:0.8, Excellent:1.0, Good:0.7, Great:0.8, High:-0.8, Low:0.8, Reasonable:0.5, Too high:-1.0, Very bad:-0.91, Very good:0.91, Very great:1.0, Very high:-1.0, Very low:1.0 |
| Camera | Amazing:1.0, Average:0.5, Awesome:1.0, Bad:-0.69, best:1.0, Better:0.8, Brilliant:0.9, Excellent:1.0, Good:0.7, Great:0.8, Nice:0.7, Not bad:0.7, Not good:-0.69, Not great:-0.4, Poor:-0.4, Very bad:-0.91, Very good:0.91, Very poor:-0.52, Worst:-1.0 |
| Battery | Amazing:1.0, Awesome:1.0, Bad:-0.69, Best:1.0, Better:0.8, Excellent:1.0, Good:0.7, Great:0.8, Nice:0.7, Not good:-0.69, Poor:-0.4, Stunning:0.9, Very bad:-0.91, Very good:0.91, Very poor:-0.52 |
| Screen | Amazing:1.0, Awesome:1.0, Bad:-0.69, Beautiful:0.85, Big:0.4, Excellent:1.0, Good:0.7, Great:0.8, Nice:0.7, Not good:-0.69, Poor:-0.4, Small:-0.4, Very poor:-0.52 |
| Phone | Amazing:1.0, Average:0.5, Awesome:1.0, Bad:-0.69, Best:1.0, Better:0.8, Brilliant:0.9, Excellent:1.0, Good:0.7, Great:0.8, Much better:0.9, Nice:0.6, Not good:-0.69, Poor:-0.4, Superb:1.0, Very bad:-0.91, Very good:0.91, Very nice:0.78, Very poor:-0.52, Worst:-1.0 |

**Table 8:** Decision Matrix $D$ for smart phones data set

| Smart Phone | Features' Score of Smart Phones | | | | |
|---|---|---|---|---|---|
| | Price | Camera | Battery | Screen | Phone |
| iPhone 7 | 0.682 | 0.752 | 0.573 | 0.592 | 0.846 |
| Samsung Galaxy S7 Edge | 0.418 | 0.746 | 0.706 | 0.624 | 0.731 |
| Google Pixel | 0.155 | 0.940 | 0.651 | 0.750 | 0.755 |
| HTC Desire 10 Pro | 0.350 | 0.483 | 0.335 | 0.403 | 0.764 |
| Lenovo Z2 Plus | 0.749 | 0.286 | 0.600 | 0.488 | 0.620 |

$$R = D \times S = \begin{bmatrix} 0.682 & 0.752 & 0.573 & 0.592 & 0.846 \\ 0.418 & 0.746 & 0.706 & 0.624 & 0.731 \\ 0.155 & 0.940 & 0.651 & 0.750 & 0.755 \\ 0.350 & 0.483 & 0.335 & 0.403 & 0.764 \\ 0.749 & 0.286 & 0.600 & 0.488 & 0.620 \end{bmatrix} \times \begin{bmatrix} 0.2613 \\ 0.0407 \\ 0.0801 \\ 0.0801 \\ 0.5377 \end{bmatrix} = \begin{bmatrix} 2.07 \\ 1.92 \\ 1.90 \\ 1.38 \\ 1.66 \end{bmatrix}$$

**Figure 10:** Steps for calculation of rank vector $R = D \times S$

**Table 9:** Smart Phones with their Rank Score

| Smart Phone | Rank | Rank Score |
|---|---|---|
| iPhone 7 | 1 | 2.07 |
| Samsung Galaxy S7 Edge | 2 | 1.92 |
| Google Pixel | 3 | 1.90 |
| HTC Desire 10 Pro | 5 | 1.38 |
| Lenovo Z2 Plus | 4 | 1.66 |

 

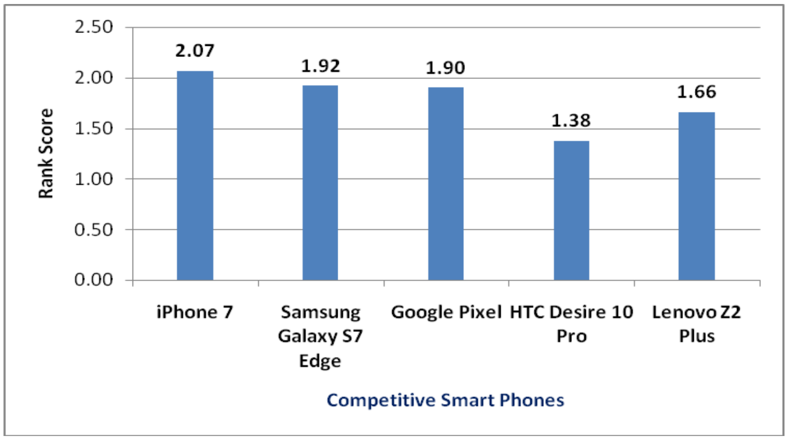


**Figure 11:** Smart Phones and their Rank Scores

**Table 10:** Rank Lists of competitive smart phones generated by our proposed method *(L)* and assigned by two domain experts *($L_1$* and *$L_2$)*

| ID | Smart Phone | Rank List generated by our proposed method (*L*) | Experts' Rank Lists | |
|---|---|---|---|---|
| | | | $L_1$ | $L_2$ |
| P1 | iPhone 7 | 1 | 2 | 1 |
| P2 | Samsung Galaxy S7 Edge | 2 | 1 | 2 |
| P3 | Google Pixel | 3 | 3 | 3 |
| P4 | HTC Desire 10 Pro | 5 | 4 | 4 |
| P5 | Lenovo Z2 Plus | 4 | 5 | 5 |

**Table 11:** Overlapping Score calculation of *L* with $L_1$ and $L_2$

| Depth (k) | A={L@k} | Overlapping between *L* and $L_1$ | | Overlapping between *L* and $L_2$ | |
|---|---|---|---|---|---|
| | | B={$L_1$@k} | \|A∩B\|/k | C={$L_2$@k} | \|A∩C\|/k |
| 1 | {P1} | {P2} | 0/1=0.00 | {P1} | 1/1=1.00 |
| 2 | {P1,P2} | {P2,P1} | 2/2=1.00 | {P1,P2} | 2/2=1.00 |
| 3 | {P1,P2,P3} | {P2,P1,P3} | 3/2=1.00 | {P1,P2,P3} | 3/3=1.00 |
| 4 | {P1,P2,P3,P5} | {P2,P1,P3,P4} | 3/4=0.75 | {P1,P2,P3,P4} | 3/4=0.75 |
| 5 | {P1,P2,P3,P5,P4} | {P2,P1,P3,P4,P5} | 5/5=1.00 | {P1,P2,P3,P4,P5} | 5/5=1.00 |
| **Average Overlap Score** | | **(0+1+1+0.75+1)/5=0.75** | | **(1+1+1+0.75+1)/5=0.95** | |
| **Aggregate Average Overlap Score** | | | **(0.75 + 0.95)/2=0.85=85.0%** | | |

## V. CONCLUSION

In this paper, we have presented opinion mining and product ranking method from customers' review documents. In opinion mining module we have used rule based method to extract the features and opinion of the product from customers reviews documents. This <*feature*, *modifier*, *opinion*> triplets are used in product ranking. The AHP first rank the identified five features. Thereafter, it ranked list of smart phones – *iPhone 7*, *Samsung Galaxy S7 Edge*, *Google Pixel*, *HTC Desire 10 Pro*, and *Lenovo Z2 Plus* using extracted triplets related to these five features. The dataset of smart phones are created by downloading from three deferent online shopping websites - *Flipkart*, *Snapdeal*, and *Amazon*. Proposed method ranked the *iPhone 7* at first with rank score *2.07* followed by *Samsung Galaxy S7 Edge* with rank score *1.92*. The evaluation result show that the proposed method gives up to marks results on ranking of competitive products.

## REFERENCES


[1] M. Hu, B. Liu, "Mining and Summarizing Customer Reviews", In the Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'04), USA, pp. 168–177, 2004.

[2] S. Kim, E. Hovy, "Determining the Sentiment of Opinions", In the Proceedings of the 20th International Conference on Computational Linguistics (COLING'04), Switzerland, pp. 1367–1373, 2004.

[3] B. He, C. Macdonald, J. He, and I. Ounis, "An Effective Statistical Approach to Blog Post Opinion Retrieval", CIKM., 10, pp. 1063-1069, 2008.

[4] B.Liu, "Sentiment Analysis and Subjectivity", "Handbook of Natural Language Processing", Second Edition, Chapman and Hall/CRC publisher, Editors (N. Indurkhya & F. J. Damerau), 2010.

[5] P. Turney, "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews", In the Proceedings of the 40th Annual Meeting on Association for Computational Linguistics (ACL'02), Philadelphia, Pennsylvania, pp. 417–424, 2002.

[6] B. Pang, L. Lee, S. Vaithyanathan, "Thumbs up? Sentiment Classification Using Machine Learning Techniques", In the Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP'02), USA, pp. 79–86, 2002.
[7] S. Arora, M. Joshi, C.P. Rose, "Identifying Types of Claims in Online Customer Reviews", In the Proceedings of the NAACL HLT., 6, pp. 37-40, 2009.
[8] B. Liu, M. Hu, J. Cheng, "Opinion Observer - Analyzing and comparing opinions on the Web", In the Proceedings of the 14th International Conference on World Wide Web (WWW'05), Japan, pp. 342–351, 2005.
[9] X. Ding, B. Liu, S.Y. Philip, "A Holistic Lexicon-Based Approach to Opinion Mining", In the Proceedings of the first ACM International Conference on Web search and Data Mining (WSDM'08), California, USA, pp. 231–240, 2008.
[10] M. Hu, B. Liu, "Mining and Summarizing Customer Reviews", In the Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'04), USA, pp. 168–177, 2004.
[11] A.M. Popescu, O. Etzioni, "Extracting Product Features and Opinions from Reviews", In the Proceedings of the 2005 Conference on Empirical Methods in Natural Language Processing (EMNLP'05), Canada, pp. 339–346, 2005.
[12] B. Liu, L. Zhang, "A survey of Opinion Mining and Sentiment Analysis", Mining Text Data, Springer, ISBN: 978-1-4614-3223-4, pp. 415 – 463, 2012.
[13] M. Abulaish, Jahiruddin, M. Doja, T. Ahmad, "Feature and Opinion Mining for Customer Review Summarization", In the Proceedings of the Pattern Recognition and Machine Intelligence (PReMI'09), pp. 219-224, 2009.
[14] W. Medhat, A. Hassan, H. Korashy, "Sentiment Analysis Algorithms and Applications: A survey", Ain Shams Engineering Journal, 5(4), pp. 1093-1113, 2014.
[15] P. Zhang, W. Yu, C. Sha, X. He, A. Zhou, "Product-Oriented Review Summarization and Scoring", Frontiers of Computer Science, 9(2), pp. 210-223, 2015.
[16] M. Chae, J. Kim, "Do size and structure matter to mobile users? an empirical study of the effects of screen size, information structure, and task complexity on user activities with standard web phones", Behaviour & Information Technology, 23 (3), pp. 165–181, 2004.
[17] C.H. Chen, L.P. Khoo, W. Yan, "Evaluation of multicultural factors from elicited customer requirements for new product development", Research in Engineering Design, 14 (3), pp. 119–130, 2003.
[18] M.C. Chuang, C.C. Chang, S.H. Hsu, "Perceptual factors underlying user preferences toward product form of mobile phones", International Journal of Industrial Ergonomics, 27 (4), pp. 247–258, 2001.
[19] S.H. Han, S.W. Hong, "A systematic approach for coupling user satisfaction with product design", Ergonomics, 46 (13-14), pp. 1441–1461, 2003.
[20] S.H. Han, K.J. Kim, M.H. Yun, S.W. Hong, J. Kim, "Identifying mobile phone design features critical to user satisfaction", Human Factors in Ergonomics & Manufacturing, 14 (1), pp. 15–29, 2004.
[21] Y.S. Lee, S.W. Hong, T.L. Smith-Jackson, M.A. Nussbaum, K. Tomioka, "Systematic evaluation methodology for cell phone user interfaces", Interacting with Computers, 18 (2), pp. 304–325, 2006.
[22] X. Yu, S. Guo, J. Guo, X. Huang, "Rank B2C e-commerce websites in e-alliance based on AHP and fuzzy TOPSIS", Expert Systems with Applications, 38 (4), pp. 3550–3557, 2011.
[23] H.Y. Hsieh, S.H. Wu, "Ranking online customer reviews with the SVR model", In the Proceedings of the 2015 IEEE 16th International Conference on Information Reuse and Integration, IEEE, pp. 550–555, 2015.
[24] J.C. Pomerol, S. Barba-Romero, "Multicriterion Decision in Management: Principles and Practice", vol. 25, 1st ed., Springer US, ISBN: ISBN 978-1-4615-4459-3, 2000.
[25] G. Isiklar, G. Buyukozkan, "Using a multi-criteria decision making approach to evaluate mobile phone alternatives", Computer Standards & Interfaces, 29 (2), pp. 265–274, 2007.
[26] S.T. Saaty, "The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation", 1st ed., McGraw-Hill, New York, 1980.
[27] R.W. Saaty, "The analytic hierarchy process-what it is and how it is used", Mathematical Modelling, 9 (3-5), pp. 161–176, 1987.
[28] W. Webber , A. Moffat , J. Zobel , "A similarity measure for indefinite rankings", ACM Transactions on Information Systems, 28 (4), pp. 20:1–20:38, 2010.

## Authors Profile

Dr. *Jahiruddin* has completed his Master degreee from Aligarh Muslim University, Aligarh and Ph.D Computer Science Degree from Jamia Millia Islamia, New Delhi. He is currently working as Assistant Professor in Department of Computer Science, Jamia Millia Islamia, New Delhi. He has published more than 15 research papers in reputed international Journa and proceedings of conferences. His main research work focuses on Text Mining, Biomedical Text Mining, Data Analytics, Graph Mining. He has more than 10 years of teaching Experience in Central University.