International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-11 Issue-3, January 2022

EAPRAST: Extensive Approach for Product Ranking in Aspect-Based Sentiment Analysis using TRIE

Nibedita Panigrahi, Asha T

Abstract: ***To assist prospective consumers make educated purchasing choices, we are analyzing and mining data from product reviews on online shopping websites. Two methods exist for extracting aspects. Rule-based and Highest Adjective Count (HAC) algorithms. The aspect ranking will use MAX opinion score method and enhanced SentiWordNet opinion score. SentiWordNet uses a hash map structure to turn keys into tiny values that may be used to index data. Hashing can search, insert, and remove in O(L) time. The disadvantage is that if two keys give the same hashCode value, the hashMap's speed suffers. When HashMap buckets are full, they need to be resized. We replaced it with TRIE, which can insert and locate strings in O(L) time, where L is the word length. TRIE is quicker than Hashing because of its implementation. Here hash function and collision handling is not required (like we do in open addressing and separate chaining). TRIE also allows us to print all words in alphabetical order, which is not feasible using hashing. TRIE can effectively search for prefixes. On the other hand, we offer a method that ranks items based on their similarity in terms of features and price. The Suggested method is applied to three conventional databases such as Amazon, Yelp, and IMDB and the solution provides a more effective and dependable online buying experience.***

Keywords: ***Sentiment Analysis, Natural Language Processing, Product Ranking, TRIE Algorithm, Hash Map***

I. INTRODUCTION

E-commerce companies and social media platforms increasingly encourage customers to publish online product reviews [1, 2]. "To help consumers make purchases, internet reviews must be ranked." The approach automatically detects the sentiment Amazon (http://www.amazon.cn/), Yelp (https://www.yelp.com/), and the Internet Movie Database (http://www.imdb.com.cn/) all allow users to post online product reviews. Online product reviews have been found to have a major influence on customer buying choices. Orientation of each online review, analyses the performance of alternative goods against each product characteristic, and ranks them. Studies on comparing items based on internet reviews are rare. To rate items based on internet reviews requires two steps: (1) evaluating online reviews and (2) rating products. The first is to discover essential product attributes or customer sentiments by analyzing internet reviews, and the second is to rate items based on the analysis findings. This research has helped rate items based on internet reviews. However, present approaches have certain drawbacks. To help the customer choose a desired product, the ranking of alternatives should be based on the attributes and weights given by the consumer. Also, research [6-11] overlook neutral sentiment orientations in reviews. This will result in lost decision information. A customer who publishes a neutral review indicates that their opinion of the product is reluctant and unclear. The information on reluctant or unsure should not be overlooked [12-14] since it helps prospective customers make informed purchases. In our suggested method (EAPRAST), we begin by obtaining reviews in certain product categories. Select 10 or more goods with comparable primary characteristics. Major characteristics are chosen after analyzing reviews and product descriptions. For TVs, we compare 10 models with the same screen size. We keep the product specifications in a database. The details include the product's features, manufacturer, description, and sales rank. The findings of our product rank technique are afterwards compared to real sales rank. Then we tag and stem the reviews. As a consequence, SA and review ranking will be important. Customers' tastes may vary according to the many elements of each product. The next step is to identify these features and weight them accordingly. In this stage, we connect sentences to attributes based on their terms. Next, we remove irrelevant material from reviews by filtering statements that don't relate to the product or its characteristics (features). Next, we give emotion values to these negative, positive, and neutral statements. The step-by-step results are utilized to rate the product (both in the general case and according to user specific preferences). We may also prioritize product features depending on customer choices. Finally, the user is given with clear search results that help them locate items quicker, more precisely, and more efficiently. Most of the time, the client does not express their decision explicitly, but rather in words that combine reviews with lines that are more general in character and have nothing to do with the product or viewpoint. The user may not directly state the features, write wrong phrases, lack punctuation,

Manuscript received on January 30, 2022.
Revised Manuscript received on February 03, 2022.
Manuscript published on February 28, 2022.
* Correspondence Author
Nibedita Panigrahi*, Assistant Professor, Department of Information Science & Engineering, RV Institute of Technology, Bangalore, (Karnataka) India. Email: kuni.kist@gmail.com, nibeditap.rvitm@rvei.edu.in
Dr. Asha. T, Professor**,** Department of Computer Science & Engineering, Bangalore, Institute of Technology, Bengaluru, (Karnataka) India. Email: asha.masthi@gmail.com, asha@bit-bangalore.edu.in.

or use grammatically inappropriate language [12]. Customer feedback impacts other consumers' purchase choices, therefore it's crucial for firms to know what customers want. The issue is that most customer evaluations are lengthy and numerous, making distillation of information difficult. A consumer will usually check a few reviews before making a purchase. It is possible that a user will make a biased choice on a product. Similarly, manufacturers want to read reviews to find out what aspects customers like or hate so they can focus on those. More significantly, the enormous quantity of reviews makes it difficult for businesses to monitor client feedback. Continue reading for more information. Section 2 summarizes research on sentiment analysis. Sec. 3 explains how to rate items using internet reviews. Summary of the problem description and suggested solution's benefits over the present system. Section 5 discusses the suggested method's two sections. In Section 6, the suggested method's experiments and results are shown. Section 7 concludes with this paper's references.

- **SENTIMENTAL ANALYSIS (SA):**

Words, phrases, and paragraphs are given polarity and sentiment values in sentiment analysis. There has been a lot of work done on sentiment analysis in recent years. There are two basic techniques in this subject. First, some works concentrate on giving a text a good or negative mood as a whole. While this works well for basic content, reviews are more difficult. The second method examines the text sentence by sentence. Our theory is that various portions of reviews discuss good and negative aspects of a product, therefore it makes sense to treat each phrase as a separate body of text. Moreover, following stages of the process need consideration of distinct product aspects. In order to analyze various parts of a review, we establish distinct approaches. We total together the sentiment values of each facet to get each review's perspective on the issue. Technically, the SA issue has two primary types. To measure 'polarity', the first technique (lexicon-based [5,30]) focuses on generating word lexicons. Polarity is the orientation of a word, phrase, or paragraph in relation to its emotion. Some works additionally save information about part-of-speech tagging to be more precise. While this strategy works well for straight and basic statements, it struggles with real language complexity like polarity. The text categorization methodology [21,33] analyses and categorizes sentence polarity as a whole. Usually, a lexicon is utilized (in some cases to be used as seed to expand and in cases as one of the classification features). Others concentrate on fragments or aspect-based SA [26]. The research suggests that classification techniques outperform lexicon-based approaches in increasingly complex texts and contexts. Before discussing our text categorization method, let us discuss the intricacy of reviews. Our purpose is to analyze reviews in terms of natural language complexity. To learn more about product reviews, we manually annotated a collection of reviews into several product categories. The goal is to see whether a basic analyzer will enough for our sentiment analysis. The findings are shown in Table 1. We classified the sentences as neutral, good, or negative. We also subdivided the positive class into three. 'This product is amazing,' for example. (2) Negation; to negate a negative in a statement, e.g., Inferring positive or negative interpretation of a statement is complex. So, using basic analysis and negation in our classifier, we can obtain close to 90% accuracy. Our study of internet reviews illustrates the inherent challenges of natural language processing. The important points are as follows: 1. A neutral statement sometimes follows a negative sentence. This neutral statement makes the prior sentence positive or vice versa. This is largely owing to our lack of semantic understanding to consider all the natural language concepts. In other words, a more sophisticated solution to this issue does not significantly enhance accuracy over the basic one. Also, the semantic information in various categories is somewhat different, which may adversely impact the outcome and nullify the process cost. Based on the aforementioned, we categorize SA traits as follows: - Structural features: they concentrate on sentence structure, e.g., negation. - Semantic features: certain words have acquired new meanings. This feature group focuses on this idea. There are SA efforts that use these terms as the main technique and manually annotate them to give values [14,20]. Most extensive methods employing this strategy involve compilations of Quirk and Crystal [25]. Similarly, we add manual weights to terms that appear more often in reviews than other portions of the literature. The final data set contains 76 words, as seen in Table 2. We begin by labelling phrases as positive or negative, from terms to full sentences. To better understand the phrases, we increase the SentiWordNet corpus [6]. SentiWordNet extends WordNet [18,19] by giving words a negative or positive value between 0 and 1. Words may be good, bad, or both. This category includes all product nouns or attributes. The usage of a word's pre-assigned negative or positive number. Words with no negative or positive valence are neutral. For the remaining three categories, we use the threshold in Eq. 1 to determine positive and negative polarity.

$$W_{pol} = \begin{cases} Positive & \text{if } W_p - W_n > \theta \\ Negative & \text{if } W_n - W_p > \theta \\ Neutral & \text{if } |W_p - W_n| < \theta \end{cases} \quad (1)$$

where Wp is the positive polarity of the word, Wn is the negative polarity and θ is our assigned threshold. Wpol holds the final polarity of the word. For example, by assigning θ as .15, for the word 'living' with the positivity of and negativity of 0.125 will result in assigning positive sentiment to the word.

Our approach to SA consists of two phases. In the first phase, we solve the problem of non-neutral terms that appear in neutral sentences.

II. RELATED WORK

Aspect-level sentiment analysis aims to determine the product's best and worst features. For those who are just interested in a few key characteristics and don't care about the rest, this may be a sensible way to let buyers weigh in on whether or not a product meets their requirements.

Figure 1 depicts the system's five primary components. Here, you'll find a breakdown of each module's functions and how they operate. Maryam K. Jawadwala et al.[1] discussed the examination of consumer feedback based on sentiment rather than specific feature. Multiple reviews are analyzed to provide a feature-based opinion summary using feature level sentiment analysis. People can make better judgments in less time if they have access to summaries of views and product features. Competitive marketing may benefit from data mining like this.

Yang Liu et al.[2] constructed an algorithm using sentiment dictionaries to determine if the reviewer is favorable, neutral, or negative about a particular characteristic of the alternative product. In order to express the performance of an alternative product in terms of a product characteristic, an intuitionistic fuzzy number is generated based on the discovered positive, neutral, and negative sentiment orientations. An IFWA operator and preference ranking organisation techniques for enrichment assessments II are used to rate alternative goods (PROMETHEE II). Yang Liu et al.[3] used the interval-valued intuitionistic fuzzy Technique for Order Preference by Similarity to an Ideal Solution for rating items based on sentiment classification and online reviews (TOPSIS).

Chonghui Guo et al.[4] offered a new ranking approach based on internet reviews that takes into account both objective and subjective sentiment evaluations for various characteristics of alternative items. To begin, the LDA topic model is used to assess the relative importance of each of these factors in determining the product's objective sentiment value. Jovelyn C. Cuizon et al.[5] forecasted the numerical ranking of text reviews in a trip diary application, we used sentiment analysis. Using the app, tourists may record and write evaluations of the places they've visited. S. K. Lakshmanaprabu et al.[6] suggested dissecting high-recommendation online business sites using a collecting technique and an improvement system has been explored. Kunpeng Zhang et al.[7] discussed use of consumer feedback to sift hundreds of thousands of features to choose the best products. To rank items based on internet reviews, one must first analyze the sentiment of reviews on similar products based on distinct attributes. Then, depending on the discovered sentiment orientations, online review rating studies may be done. Thus, related research may be divided into two categories: studies on sentiment analysis and studies on rating items based on online reviews. A short literature review is offered for each area. Analyzing sentiment analysis and opinion mining are related research topics that study how people express themselves in texts [18-21]. Many applications exist for sentiment analysis and opinion mining. Companies may undertake market research and uncover weaknesses of goods and services [23-25], and customers can make better purchase choices [11, 12, 24]. In some research, sentiment analysis and opinion mining are considered interchangeable [18-21]. However, some academics believe sentiment analysis and opinion mining are distinct [18, 26]. As opposed to sentiment analysis, opinion mining focuses on finding the sentiment represented in a text before evaluating it [18, 26]. We specialize on sentiment analysis since online reviews are frequently ranked depending on their sentiment. Some researchers are interested in sentiment analysis. More than a dozen sentiment analysis approaches have been presented recently [18-21]. The classification framework of present sentiment analysis algorithms may be described by Fig. 1 [18, 20, 27]. As shown in Fig. 1, present sentiment analysis algorithms fall into two categories: machine learning-based [28-37] and lexicon-based [25, 38-41]. On one side, machine learning-based sentiment analysis algorithms may be separated into three subclasses:

- sentiment analysis techniques based on supervised machine learning [28-32], sentiment analysis techniques based on unsupervised machine learning [33-35], and Sentiment analysis techniques based on semi-supervised machine learning [36,37].

A collection of labelled training samples is required for supervised machine learning approaches [28-32]. Using the data, one may train decision tree classifiers, linear classifiers, rule-based classifiers, and probabilistic classifiers [28- 32]. In the absence of labelled training examples, unsupervised machine learning approaches may be applied [33- 35]. "Unsupervised machine learning algorithms analyze text similarities first using keyword lists of categories, then grouping texts into numerous groups based on text similarities [33-35]." Semi-supervised learning approaches may be utilized for partly labelled data, such as partially phrases or partially features [36, 37]. The lexicon-based sentiment analysis approaches are further separated into two subclasses: (1) dictionary-based [38-40] and (2) corpus- based [25, 41]. Constructing sentiment dictionaries is the cornerstone of dictionary-based sentiment analysis. Manually choose a limited group of sentiment terms for dictionary-based sentiment analysis. Then, in well-known corpora like WordNet, HowNet, or Thesaurus, new emotion words are identified and the number of sentiment words increases. Thereafter, the emotion dictionaries are determined based on the acquired sentiment words [38-40]. The difficulty of discovering opinion words with context relevant orientations is solved via corpus-based sentiment analysis. To identify syntactic patterns and a seed list of opinion words. On uses statistical approaches to uncover co-occurrence patterns or seed opinion words, whereas semantic techniques are utilized to derive semantic meanings for words based on similarity [25, 41].To conserve space, these sentiment analysis approaches are not described in depth. Literatures provide extensive reviews of previous investigations [18-21]. Zhang et al.[8] focused on rating items using internet reviews and suggested a directed and weighted product graph. The approach distinguishes between subjective and comparative language in online product evaluations.

Then, using sentiment analysis, a directed and weighted product graph is created, reflecting both subjective views and comparison relationships of the items. Finally, a new page-rank method based on the directed and weighted product network is developed. On this premise, Zhang et al.[9] suggested a system for rating items based on internet reviews.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

"The approach divides internet reviews into subgroups depending on the product characteristic stated in the sentences." To rate goods, we apply the Zhang et al.[8] approach. Then, Zhang et al.[10] used the amount of helpful votes obtained by each review to determine the age relevance of each review in determining the ranking of items.

III. PROBLEM DEFINITION

An opinion lexicon built from the WordNet database, which assigns a numerical score to each phrase, indicates whether it elicits positive or negative feelings. For sentiment classification tasks, SentiWordNet seems to be a valuable resource. However, SentiWordNet with hash-map gives the following restrictions.

- Misspellings and grammatical mistakes may cause the analysis to overlook important words or usage.
- Sarcasm and irony may be misinterpreted.
- Analysis is language-specific.
- Structure is hash-mapped which slows down the operations.
- Discriminating jargon, nomenclature, memes, or turns of phrase may not be recognize

1. Advantages of TRIE algorithm over Hash-map technique

- TRIE is not necessarily in in-order.
- Different signatures will produce different TRIEs.
- We did not compare with key in node.

2. Varieties of TRIEs

- Simple TRIE: 'store keys in interior nodes.
- Full TRIE: 'store keys in leaves'
- Compressed TRIE: 'store keys in leaves and compress paths.
- Patricia TRIE: 'combination of simple TRIE and compressed-TRIE'.

3. Properties of TRIE

- The insert and the search algorithm have the best time complexity, i.e., O(n), faster than even the best of BST.
- All words can be easily printed in alphabetical order, which is difficult if we use hashing.
- Prefix search is easily doable.
- For certain use cases that seek a higher level of accuracy, it may be worth evaluating alternatives.
- Here ,we have implemented the SentiWordNet with TRIE algorithm which is having a fast look-up table.

IV. METHODOLOGY

The main contributions in this work are:

➢ Using NLP and dynamic programming approaches, we can recognize subjective/comparative words (product aspect-based) in reviews and estimate their sentiment orientations.

➢ It is possible to create a graph that accurately depicts the intrinsic quality of a product by using sentence classification algorithms. For example, a product rating system that utilizes this enormous network to score each product aspect. In other words, the algorithm generates a rating list that a prospective buyer may use to identify the best items based on the relevance of one or more product qualities. There are three types of opinion mining approaches [8].

➢ Aspect level: In this method, the specific components of a product are categorized and the comments or reviews are taken for those aspects individually.

➢ Sentence level: The opinions expressed here are subjective. This method has the advantage of allowing customers to learn about a wide range of various sorts of consumer feedback. This method focuses on the distinction between subjective and objective data. Negative or positive opinions are based on subjective knowledge, whereas objective information is simply the truth.

➢ Document level: One individual is responsible for the whole paper in this case, which was created only for the product. As a result, the buyer will only learn about the opinion of one other person.

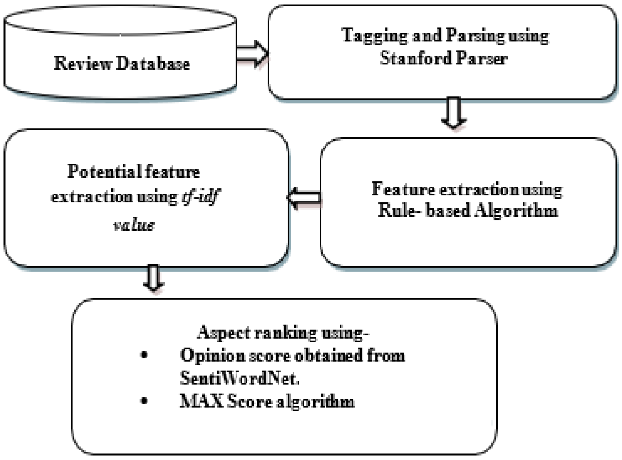

Fig-1: Architecture of the proposed Sentimental Mining system

Feature Extraction using Rule-based Algorithm: It is the goal of feature-level sentiment analysis to discover the excellent and unfavorable characteristics of a particular product feature-wise. For those who are just interested in a few key aspects and don't care about the rest, this may be a sensible way to let buyers weigh in on whether a product meets their requirements. [1] Figure 1 depicts the system's five primary components. Here, you'll find a breakdown of each module's functions and how they operate.

$$tf-idf(t_i) = tf(t_i) \times idf(t_i) \qquad 1.$$

$$idf(t_i) = \log\left(\frac{|D|}{|\{d_j : t_i \in d_j\}|}\right) \qquad 2.$$

Where, *tf*(ti) is the number of documents containing ti |D| is the total number of documents.

Relevant features are all noun phrases with a tf-idf value greater than a certain threshold.

Next, a list of all views and modifiers is generated for each product feature, which is then utilized to determine the polarity of opinion phrases.

The most common method of retrieving values from a data structure is to utilize hash tables. TRIEs are far more efficient than hash tables and have various benefits over the same, despite the fact that they are less well-known than hash tables. Mainly:

- There won't be any collisions hence making the worst performance better than a hash table that is not implemented properly.
- No need for hash functions.
- Lookup time for a string in TRIE is O(k) where k = length of the word.
- It can take even less than O(k) time when the word is not there in a TRIE.

```
ALGORITHM: INSERT (KEY, VALUE)
Input: key-value pair
if TRIE is empty
  root = create new root with key-value pair;
  return
endif
  // Start numbering the bits from 0.
recursiveInsert (root, key, value, 0)

Algorithm: recursiveInsert (node, key, value, bitNum) Input: TRIE
node, key-value pair, which bit we are using now
  // Compare with node key to see if it's a duplicate.
if node.key = key
  Handle duplicates;
      return
endif
// Otherwise, examine the bitNum-th bit
if key.getBit (bitNum) = 0
  // Go left if possible, or insert.
  if node.left is null
    node.left = new TRIE node with key-value;
  else
    // Note: at next level we'll need to examine the next bit.
    recursiveInsert (node.left, key, value, bitNum+1)
  endif
else
  // Same thing on the right
  if node.right is null
    node.right = new TRIE node with key-value;
  else
    recursiveInsert (node.right, key, value, bitNum+1)
  endif
endif
```

Search:

- Search is straightforward.
- Compare the input key with the current node.
- If equal, the key is found; return.
- Otherwise, examine *i*-th bit (at level *i*) and go left or right accordingly.
- If next link is null, search ends without finding the key.

```
Pseudocode:
Algorithm: search (key)
Input: search-key
node = recursiveSearch (root, key, 0)
if node is null
  return null
else
  return node.value
endif
Output: value, if key is found
```

```
Algorithm: recursiveSearch (node, key, bitNum)
Input: TRIE node, search-key, which bit to examine
// Compare with key in node.
if node.key = key
  return node
endif
// Otherwise, navigate further.
if key.getBit (bitNum) = 0
  if node.left is null
    // Not found => search ends.
    return null
  else
    // Search left.
    return recursiveSearch (node.left, key, bitNum+1)
  endif
else
  if node.right is null
    // Not found => search ends.
    return null
  else
    // Search right.
    return recursiveSearch (node.right, key, bitNum+1)
  endif
endif
Output: TRIE node if found, else null.
```

Feature Extraction using improved HAC Algorithm: The High Adjective Count algorithm (HAC) [3] is another way for extracting aspects. The approach is based on the premise that nouns with a high number of reviews are more likely to have meaningful and distinctive characteristics than those with less reviews. The opinion scores given in the modified HAC are used to rank the retrieved aspects using the Max Opinion Score method. Semantic connections connect opinion words and product aspects in Rule-based Algorithm; therefore, each phrase is transformed to Stanford Dependency (SD) using Stanford parser [11].

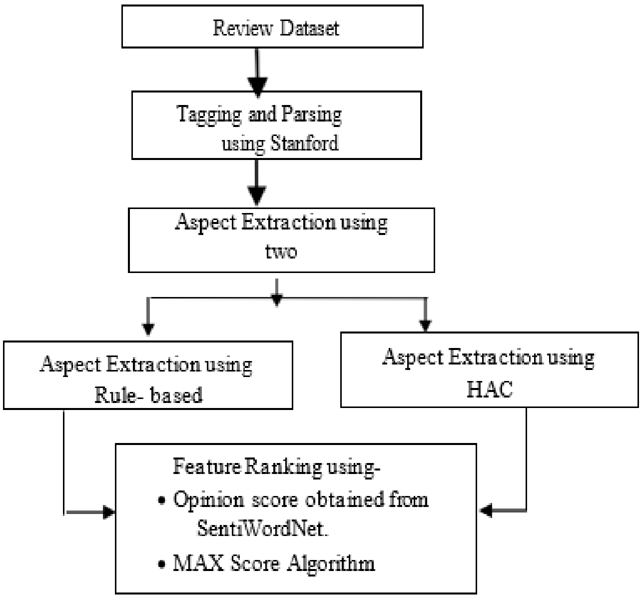

Fig-2: Block diagram for system Implementation.

1. Hotel is awesome.
2. Although hotel is far from city, Food quality is not good.
3. Food taste is fantastic.
4. Quality of sound is the best.
5. Good screen size with low cost.
6. Battery life is satisfactory as compared to another android phone.
7. Touch screen is not very impressive.
8. Quality of sound is really good.
9. Camera zoom is not good.
10. Battery life is poor.
11. Touch screen is wonderful.
12. Quality of sound is not good.
13. A decent phone.
14. Low-cost phone.
15. Camera does not capture clear picture during night.

Fig. 3: Sample reviews of IMDB.

Food Quality is awesome.
Food _NN quality_NN is_VBZ awesome_JJ ._.
nn(quality-2, Food -1)
nsubj(awesome-4,quality-2)
cop(awesome-4, is-3)
root(ROOT-0, awesome-4)

Bag color is fantastic.
Bag_NN color_NN is_VBZ fantastic_JJ ._.
nn(color-2, Bag-1)
nsubj(fantastic-4, color-2)
cop (fantastic-4, is-3)
root(ROOT-0, fantastic-4)

Fig. 4: Stanford Dependency (SD) for partial reviews of figure 3.

Aspect Ranking: Aspect ranking is performed using following steps The polarity of extracted opinions for each feature is classified using

1) Opinion score obtained from SentiWordNet [6].
2) Assign polarity manually in the range of [-4, 4], known as MaxOpinionScore algorithm.

The overall weight of a feature is calculated by multiplying the polarity value of the opinion word with the number of sentences which contain that opinion. This is mentioned in equation 3.

$$TotalWt = \sum_{n=1}^{d}(WPF - WNF) \qquad 3.$$

WPF = Weight of positive features.
WNF = Weight of negative features.

V. RESULT AND DISCUSSION

For the three standard datasets (Amazon, Yelp, and IMDB), we've developed the Rule-based algorithm and picked the highest score. Figure 6 depicts a snapshot of the customer reviews, showing the rating and the percentage of positive to negative feedback. Using SentiWordNet and a hash map, Table-1 compares rule-based algorithm rankings. Finally, in Table-2, a comparative analysis of rule-based algorithm ranking utilizing SentiWordNet and TRIE algorithm is shown. Table 3 demonstrates the comparison of the HAC algorithm for rating products using SentiWordNet and Hash-map .'s As a result, SentiWordNet with Hash-map is used in Rule-based method in Table-4. "When employing SentiWordNet with a TRIE-algorithm, performance metrics are shown in Table-4." Rule-based approach combining SentiWordNet with Hash-map is shown in Fig.

Finally, Fig exhibits graph metrics for the Rule- based method employing SentiWordNet and TRIE-algorithm.

So there is no way for me to plug it in here in the US unless I go by a converter. 1
Negative
Very Negative Count: 0 Negative Count: 1.
Neutral Count: 0.
Very Positive Count: 0 Positive Count:
0. MAX SENTIMENT COUNT IS 1
****OVERALL SENTIMENT IS NEGATIVE**** 0
Good case, Excellent
value. 0
Good case, Excellent
value. 3
Positive
Very Negative Count: 0 Negative Count: 1.
Neutral Count: 0.
Very Positive Count: 0 Positive Count:
1. MAX SENTIMENT COUNT IS 1
****OVERALL SENTIMENT IS NEGATIVE****
I have two more years left in this contract and I hate this phone. 3
Positive
Very Negative Count: 0 Negative Count: 8.
Neutral Count: 11.
Very Positive Count: 1 Positive Count:
16. MAX SENTIMENT COUNT IS
16
****OVERALL SENTIMENT IS POSITIVE****

Fig 5: Screenshot of Ranking for the amazon dataset for the Rule-based algorithm.

Retrieval Number: 100.1/ijitee.C97620211322
DOI: 10.35940/ijitee.C9762.0111322
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

Table 1: Comparative study of Rule-based algorithm ranking using SentiWordNet with Hash-map

SL No.	Product Name	Subjective sentences only (Score using SentiWordNet)			
		Positive Features	Rank	Negative Features	Rank
1	Apple-iphone	Quality	6.5	Lens unit	-6.5
2	Mi-note 10	resolution	3.8	period	-3.5
3	Samsung–galaxy M32	lens	11	shooting	-11
4	Vivo v21	Magnesium finish	9	strap	-4

Table 2: Comparative study of Rule-based algorithm ranking using SentiWordNet algorithm with TRIE

SL No.	Product Name	Subjective sentences only (Score using SentiWordNet)			
		Positive Features	Rank	Negative Features	Rank
1	Apple-iphone	Quality	6.5	Lens unit	-6.5
2	Mi-note 10	resolution	3.8	period	-3.5
3	Samsung–galaxy M32	lens	11	shooting	-11
4	Vivo v21	Magnesium finish	9	strap	-4

Table 3: Comparative study of HAC algorithm for product ranking

SL No.	Product Name	Subjective sentences only (Score using SentiWordNet)			
		Positive Features	Rank	Negative Features	Rank
1	Apple-iphone	Quality	6	Lens unit	-5.5
2	Mi-note 10	resolution	7	period	-7
3	Samsung–galaxy M32	lens	11	shooting	-7
4	Vivo v21	Magnesium finish	7	strap	-6.5

Table 4: Performance measures for Rule-based algorithm using SentiWordNet with Hash-map

Product	Aspects	Precision	Recall	F1-measure
Apple-iphone	Quality	76	45	50
Mi-note 10	resolution	77	100	60
Samsung–galaxy M32	lens	73	60	50
Vivo v21	Magnesium finish	74	50	65

Table 5: Performance measures for Rule-based algorithm using SentiWordNet with TRIE

Product	Aspects	Precision	Recall	F1-measure
Apple-iphone	Quality	81	50	50
Mi-note 10	resolution	79	100	61
Samsung–galaxy M32	lens	74	60	57
Vivo v21	Magnesium finish	72	55	65

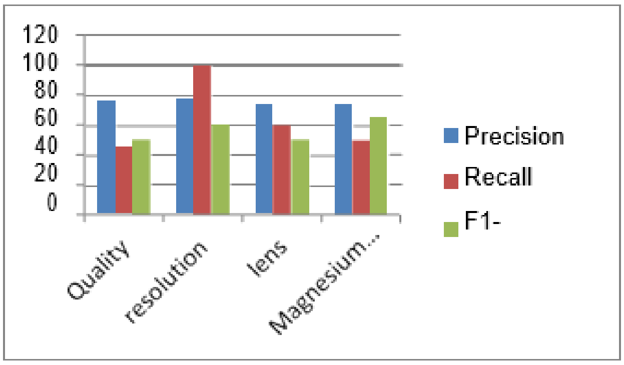

Fig 6: Graph measures for Rule-based algorithm using SentiWordNet with Hash-map.

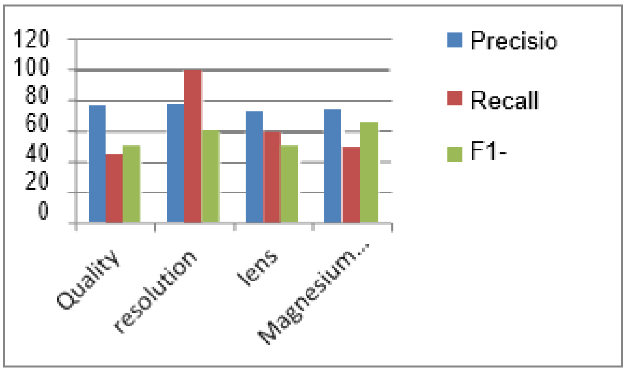

Fig 7: Graph measures for Rule-based algorithm using SentiWordNet with TRIE

ACKNOWLEDGMENT

This research was supported by Bangalore Institute of Technology. We want to thank everyone who supported us and motivated us during the course of the research. We also show the gratitude to everyone who reviewed and suggested insightful feedbacks. At last, we also want to motivate other readers of the work to carry out further research in numerous fields of cyber security and make Internet a safer & better place for everyone.

VI. CONCLUSION

As a solution to the issue of product rating, we provide a new technique in this study. To the best of our knowledge, this is the first time the SentiWordNet decision-making process has taken the TRIE algorithm into account. It is also an innovative way to enhance the ranking process for new goods by using TRIE algorithm in the area of information retrieval. For the product ranking, we have also examined the HAC method and implemented it on three common datasets using the HAC algorithm. With limited resources, our studies and findings constitute a successful proof of concept.

REFERENCES

1. Maryam K. Jawadwala and Seema Kolkur, ‘Feature Ranking in Sentiment Analysis’, International Journal of Computer Applications (0975 – 8887) Volume 94 – No 13, May 2014.

2. Yang Liu, Jian-Wu Bi, Zhi-Ping ,'Ranking products through online reviews: A method based on sentiment analysis technique and intuitionistic fuzzy set theory', Information Fusion, Northeastern University, Shenyang, China(2016), pp. 121–128.
3. Yang Liu, Jian-Wu Bi and Zhi-Ping Fan, 'A Method for Ranking Products Through Online Reviews Based on Sentiment Classification and Interval-Valued Intuitionistic Fuzzy TOPSIS',International Journal of Information Technology & Decision Making, China(2017), Vol. 16, No. 06, pp. 1497-1522 .
4. Chonghui Guo Zhonglian Du Xinyue Kou, 'Products Ranking Through Aspect-Based Sentiment Analysis Of Online Heterogeneous Reviews', Journal of Engineering Society of China, (2018), Vol. 27, No.5,pp: 542-558.
5. Jovelyn C. Cuizon Carlos Giovanni Agravante,'Sentiment Analysis for Review Rating Prediction in a Travel Journal', Association for Computing Machinery, NLPIR Korea(2020),pp. 1-9.
6. K S. K. Lakshmanaprabu and K. Shankar , 'Ranking Analysis for Online Customer Reviews of Products Using Opinion Mining with Clustering', research article in Graduate Program in Applied Informatics, University of Fortaleza, Fortaleza, CE, Brazil,Vol. 2018,pp. 1-9.
7. Kunpeng Zhang and Ramanathan Narayanan,' Voice of the Customers: Mining Online Customer Reviews for Product Feature-based Ranking', in Proc. 3rd Conf. Online Social Networks ,Northwestern University 2145 Sheridan road, Evanston, IL, 60208, USA, (2015), pp. 1–8.
8. Zhang, R. Narayanan and A. Choudhary, 'Mining online customer reviews for ranking products', EECS Department, Northwestern University (2009).
9. Zhang, R. Narayanan and A. N. Choudhary, Voice of the customers: Mining online customer reviews for product feature-based ranking, in Proc. 3rd Conf. Online Social Networks, Boston, MA, USA (2010), pp. 1–9.
10. K. Zhang, Y. Cheng, W. K. Liao and A. Choudhary, Mining millions of reviews: A technique to rank products based on importance of reviews, in Proc. 13th Int. Conf. Electronic Commerce, Liverpool, United Kingdom (2011), pp. 121–128.
11. K. Chen, G. Kou, J. Shang, Y. Chen, Visualizing market structure through online product reviews: Integrate topic modeling, TOPSIS, and multi-dimensional scaling approaches, Electron. Commer. Res. Appl. 14 (2015),pp.58–74.
12. E. Najmi, K. Hashmi, Z. Malik, A. Rezgui, H.U. Khan, CAPRA: A comprehensive approach to product ranking using customer reviews, Computing, 97 (8) (2015),pp. 843–866.
13. M. Hu and B. Liu, Mining and Summarizing Customer Reviews, Proceedings of the 10th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD-2004), 8 (2004), pp. 168–174.
14. M. Hu and B. Liu, Mining Opinion Features in Customer Reviews, Proceedings of the 19th National Conference on Artificial Intelligence., 7 (2004), pp. 755-760.
15. Popescu and O. Etzioni, Extracting product features and opinions from reviews, Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing., (2005), pp. 339- 346.
16. Liu, M. Hu, and J. Cheng, Opinion Observer: Analyzing and Comparing Opinions, WWW., 5 (2005), pp. 342-351.
17. W. Medhat, A. Hassan, H. Korashy, Sentiment analysis algorithms and applications: A survey, Ain Shams Eng. J. 5 (2014),pp. 1093-1113.
18. K. Ravi, V. Ravi, A survey on opinion mining and sentiment analysis: Tasks, approaches and applications, Knowl.-Based Syst. 89 (2015),pp.14-46.
19. J. Serrano-Guerrero, J.A. Olivas, F.P. Romero, E Herrera-Viedma, Sentiment analysis: A review and comparative analysis of web services, Inform. Sci. 311 (2015),pp. 18-38.
20. J. A. Balazs, J. D. Velásquez, Opinion Mining and Information Fusion: A survey, Inform. Fusion 27 (2016),pp. 95–110.
21. J. C. Bertot, P. T. Jaeger, D. Hansen, The impact of polices on government social media usage: Issues, challenges, and recommendations, Gov. Inform. Q. 29(1) (2012), pp. 30-40.
22. O. Netzer, R. Feldman, J. Goldenberg, M. Fresko, Mine your own business: Market-structure surveillance through text mining, Market. Sci. 31 (3) (2012), pp. 521-543.
23. E. Marrese-Taylor, J.D. Velásquez, F. Bravo-Marquez, A novel deterministic approach for aspect-based opinion mining in tourism products reviews, Expert Syst. Appl. 41 (2012),pp. 7764–7775.
24. W.H. Zhang, H. Xu, W. Wan, Weakness Finder: Find product weakness from Chinese reviews by using aspects based sentiment analysis, Expert Syst. Appl. 39 (11) (2012) ,pp. 10283–10291.
25. M. Tsytsarau, T. Palpanas, Survey on mining subjective data on the web, Data Min. Knowl. Discov. 24 (2012,pp. 478–514.
26. F.H. Khan, U. Qamar, S. Bashir, eSAP: A decision support framework for enhanced sentiment analysis and polarity classification, Inform. Sci. 367-368 (2016),pp. 862-873.
27. X. Bai, Predicting consumer sentiments from online text, Decis. Support Syst. 50 (2011) 732–42.
28. H. Kang, S.J. Yoo, D. Han, Senti-lexicon and improved Naïve Bayes algorithms for sentiment analysis of restaurant reviews, Expert Syst. Appl. 39 (2012),pp. 6000–6010.
29. Duric, F. Song, Feature selection for sentiment analysis based on content and syntax models, Decis. Support Syst. 53 (2012),pp. 704–711.
30. F. Tian, F. Wu, K.M. Chao, Q.H. Zheng, N. Shahc, T. Lan, J. Yue, A topic sentence-based instance transfer method for imbalanced sentiment classification of Chinese product.
31. D.W. Zhang, H. Xu, Z.C. Su, Y.F. Xu, Chinese comments sentiment classification based on word2vec and SVM perf, Expert Syst. Appl. 42 (4) (2015),pp. 1857–1863.
32. Y. Li, Q. Ye, Z. Zhang, T. Wang, Snippet-based unsupervised approach for sentiment classification of Chinese online reviews. Int. J. Inf. Technol. Decis. Making, 10 (6) (2011), pp.1097-1110.
33. P. Turney, Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews, In: Proceedings of annual meeting of the Association for Computational Linguistics (2002).
34. F.H. Khan, S. Bashir, U. Qamar, TOM: twitter opinion mining framework using hybrid classification scheme, Decis. Support Syst. 57 (2014),pp. 245–257.
35. F. H. Khan, U. Qamar, S. Bashir, SWIMS: Semi-supervised subjective feature weighting and intelligent model selection for sentiment analysis, Knowl.-Based Syst. 100 (2016), pp. 97-111.
36. F.H. Khan, U. Qamar, S. Bashir, Building normalized SentiMI to enhance semi-supervised sentiment analysis, J. Intell. Fuzzy Syst. 29 (2015),pp. 1805–1816.
37. M. Hu, B. Liu, Mining and summarizing customer reviews. In: Proceedings of ACM SIGKDD international conference on Knowledge Discovery and Data Mining (2004).
38. S. Kim, E. Hovy, Determining the sentiment of opinions. In: Proceedings of interntional conference on Computational Linguistics (2004).
39. Moreo, M. Romero, J.L. Castro, J.M. Zurita, Lexicon-based comments-oriented news sentiment analyzer system, Expert Syst. Appl. 39 (10) (2012) ,pp. 9166–9180.
40. K. Xu, S.S. Liao, J. Li, Y. Song, Mining comparative opinions from customer reviews for competitive intelligence. Decis. Support Syst. 50 (2011) ,pp. 743–754.

AUTHORS PROFILE

Nibedita Panigrahi, working as Assistant Professor in the Department of Inf ormation Science & Engineering, RV Institute of Technology, Bangalore. Currently pursuing Ph.D in Computer Science & Engineering at VTU in the area of Data Mining, published 5 papers in International journals and Conferences.

Dr. Asha. T, is a Professor in the Department of Computer Science & Engineering, Bangalore Institute of Technology, Bengaluru. She obtained her Ph.D in Com puter and Information Science from Visvesvaraya Technological University, Karnataka. She has published around 31 papers in International/National journals and Conferences. Her research interests include Data Mining, Medical Informatics, Machine Learning, Pattern Recognition and Big data management etc.