IOSR Journal of Engineering (IOSRJEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
Vol. 09, Issue 11, November. 2019, Series -II, PP 10-18

Analysis and Prediction of Early Reviewers and Product Popularity Using Discriminate Novel Embedding-Based Ranking (DNE-BR)

Arifa.P[1], Ms. D. Priyadarshini[2]
M.Phil Scholar, Department of Computer Science, Sree Narayana Guru College, Coimbatore, Tamil Nadu, India[1]
Assistant Professor, Department of Computer Science, Sree Narayana Guru College, Coimbatore, Tamil Nadu India[2]
Received 29 October 2019; Accepted 15 November 2019

Abstract: E-commerce websites become very popular and tremendous growth in product marketing. Customer satisfaction and effective fulfillment of customer needs are more important in e-commerce applications. To achieve effective results and feedback in e-commerce applications, the need for effectively gathering user reviews and feedback is needed. However, the review collection and analysis became very challenging due to its size and unique nature early review prediction from the posted reviews in real-time large scale e-commerce websites are more challenging. So, the utilization of data mining will resolve these issues. There are numerous approaches and techniques proposed to handle these issues. In this proposal, Discriminate Novel Embedding-Based Ranking (DNE-BR) is proposed for reviewer prediction and product popularity detection. This work enhances the existing margin based embedding ranking model (MERM). The proposed work effectively overcomes the problem of studying the user behavior from their online product reviews. This is generally a cold start problem. The reviewers are classified into three classes namely primitive, popular and modern. This is suitable for both product and reviewer classification. Here the primitive users are the early reviewer, popular is the most frequent reviewer and the modern is the new reviewer. This classification is performed and correlated among other reviewers and their interest. The proposed work can also determine the product popularity and the product usage by different customers from online shopping websites. This kind of analysis on product reviews and reviewers also indicates that early reviewers' ratings and their received product are likely to induce product popularity. By viewing review posting process as a multiplayer competition game, we propose an improved version of margin-based embedding model for early reviewer prediction namely DNE-BR. Extensive experiments on different e-commerce datasets have shown that the proposed approach outperforms the MERM technique.

Keywords: Early Reviewers, Analysis, Prediction, MERM technique.

I. INTRODUCTION

In the present world each and every people are widely use an internet. The internet users have been increasing every day. Online shopping trend increases as internet usage increases. Online consumer reviews influence the consumer decision-making. End-user has seen the reviews of the product of the previous user and decides about good things and bad things. The internet provides an extensive source of consumer reviews, but one can hardly read all reviews to obtain a fair evaluation of a product or service. On the basis of this previous theory the process of computationally identifying and categorizing opinions expressed in a piece of text, especially in order to determine whether the writer's attitude towards a particular topic, product, etc. is positive, negative, or neutral. So, in this paper we are working on the sentiment analysis of that particular review and gives proper recommendation to end user. We are work on the supervised and unsupervised methodology. This system uses the real-time dataset of the review of the product.

Currently if we want to purchase something, we go online and search for products and look for their reviews. A user has to go through each and every review for getting information regarding each and every aspect of product. Some of these reviews contains large amount of text and detailed information about product and its aspects. A user may have to go through all of these reviews for help in decision making. Some of these products can have large amount of reviews and can contain information about its aspects in the form of large texts corpuses. A user might get irritated while reading all of these review and learn about the product. To avoid this, a system is needed that can analyze these reviews and detect the sentiments from these reviews for every aspect. In previous approaches fails to cover the fact if two reviews are mentioning same aspect with two

different words. A previous system considers those as two different aspects. Also, the aspect wise information is not preserved by these systems as they rely mostly on rating that is provided by different users for showing the quality or overall rating. The paper proposes a system that can use this information from reviews to evaluate the quality of these products' aspects. Also, the proposed system categorizes these aspects so that problem with different words for same aspects can be resolved. These aspects are identified using supervised and unsupervised techniques. Then these identified aspects are categorized in categories. The sentiments or opinions user provided for particular aspect is assigned to category of that aspect. These ratings are used to evaluate the quality of the products.

II. RELATED WORK

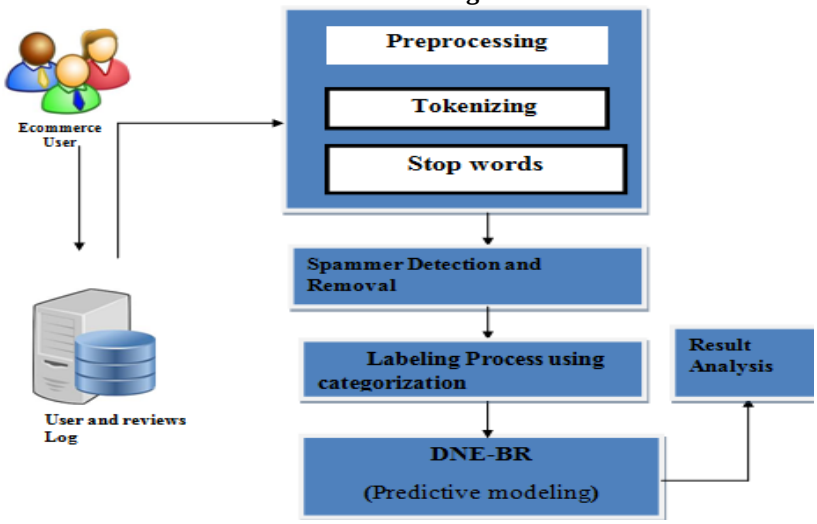
In this paper [1] the creator portrays online client audits are a feasible type of informal which assume an undeniably significant job in internet business. Through, low quality audits can frequently make issues survey watchers. [2] The creator portrays Product audits are posted online by the hundreds and thousands for well known items. Overseeing such a high volume of much of the time created online substance is a difficult errand for buyer, merchants and specialists. [3] the creator depicts as the prevalence of freestyle client produced surveys in internet business and audit sites keeps on expanding, there is a developing requirement for programmed methods that different through the tremendous number of audits and perceive quality substance. [4] The creator portrays online audit is a significant type of electronic informal that enables clients to settle on buying choices. In a lot of audits, the survey with the most supportiveness votes is viewed as generally accommodating.

In this paper [5] the creator portrays online audits given by buyers are a significant resource for internet business stages, impacting potential customers in settling on buying choices. [6] The author describes online audits are frequently our first port of call when considering items and buys on the web. [7] The author describes Articles posted on a discussion regularly contain new Internet words identified with sentiment components (highlight words and assessment words). [8] The author describes Programming improvement is exceptionally reliant on human endeavors and joint efforts, which are colossally influenced by feelings. This paper introduces a quantitative exact investigation of the enthusiastic varieties in various sorts of advancement exercises e.g., bug-fixing undertakings, improvement periods i.e., days and times and in ventures of various sizes including groups of variation sizes. [9] The author gave Collaborative sifting has demonstrated to be important for prescribing things in a wide range of spaces. [10] The author as breaking news unfurls individuals progressively depends via web-based networking media to remain side by side of the most recent updates. The utilization of internet based life in such circumstances accompanies the admonition that new data being discharged piecemeal may energize bits of gossip, a significant number of which stay unconfirmed long after their place of discharge. [11] The author describes this paper proposes a way to deal with examine consumer loyalty from Chinese audits utilizing supposition mining. Conclusion mining innovation is received to dissect consumer loyalty and find valuable data that can give choice help. [12] the author describes Supposition Mining is a bringing research field of up interest, with its various applications determined by market needs to break down item audits or to survey the general conclusion, for political reasons, during presidential battles. [13] The author describes Viewpoint extraction is one of most testing errands in sentiment mining. Numerous examines have endeavored to tackle this issue for English content. [14] the author describes With the quickly expanding number of Thai online client audits accessible in web based life and sites, conclusion examination method, additionally called sentiment mining, has turned into a significant assignment before hardly any years. [15] the author describes Feeling mining is a specialty of following the state of mind of the general population about a specific item or point from a gigantic arrangement of suppositions or audits publically accessible in web.

III. PROPOSED WORK

This section entirely discusses regarding the research system methodology and the complete process involved in this proposed research work. The system design a new effective Early reviewer framework called as **Discriminate Novel Embedding-Based Ranking (DNE-BR)** this finds the Early Reviewers.

Architecture Diagram



The proposed system focus on first formally defines the early reviewer prediction task and after that apply DNE-BR framework for effectively analysis, Early Reviewers to monitor and manage early promotion.

3.1 Contribution of the Proposed Work

- The followings are the important contributions work of the proposed research system.
- The existing framework MERM has cold-start problem this failed to provide accurate Early Reviewers.
- The proposed new framework DNE-BR Discriminate Novel Embedding-Based Ranking to find detect Early Reviewers accurately.
- The proposed system use buying behavior analysis model to identify and remove Spammer review.
- This system quantitatively analyzes the characteristics of early reviewers and their impact on product popularity.
- Real-world product review data is often imperfect and inconsistent and is likely to contain many errors. Proposed we apply Data preprocessing for resolving such such types of issues.
- The proposed apply categorization method to automatically assign each review into one of the three categories.

IV. RESEARCH METHODOLOGY

The chapter from top to bottom discusses about the algorithm and technique incorporated in the Proposed research work. Here with list of following methodology explains in this chapter one by one.

1. **Data set collection**
2. **preprocessing process**
3. **Spammer Detection and Removal**
4. **Dynamic Labeling using categorization**
5. **DNE-BR(Predictive modeling)**
6. **Result Analysis**

a) **Feature Selection**

To further decrease the number of words that should be used also indexing or keyword selection algorithms can be used. In this case, only the selected keywords are used to describe the documents

Input: sentence or document Q
Output: Pattern Pn
Matched string S
Steps:

1. Split the doc Q into number of pattern P
2. Given a set of patterns, P1, P2, ..., Pn ,
3. Give input document text T
4. find all occurrences of P in a text T = b1b2...bm.
5. do
6. if (text letter == pattern letter)
7. compare next letter of pattern to next
8. letter of text
9. else
10. move pattern down text by one letter
11. while (entire pattern found or end of text)
12. end

b) **User Interest Factor**

The similarity value calculation between users in the same category by means of interpersonal and intra personal as well as product based is important task in P-RPS.

The system proposed the effectiveness of P_RPS model with consideration of individual preference, interpersonal influence and intra personal influence. The system considers the independence of user interest in the e shopping domain. It means this can recommend items based on user interest at a certain extent this also utilizes user's association with the items to train the hidden feature vectors in stacking algorithm, especially for the existing and more expecting users.

The system also considers the Interest circle inference technique. As per the stacking algorithm this segments the social network into several sub-networks and each of them correspond to particular item collection. To overcome the cold start users who has a few rating records from the ulog then the ratings of their associated user's interest in the same category to link user interest products.

Individual Interest measure:

Due to the individuality especially users with huge rating records in e shopping domain, users usually choose products all by themselves with little influence others. To provide the product recommendation without affecting individuality for experienced users, the system proposed an optimal personalized recommendation system. The significance of user and item depends on the relevance of user interest *Tu* and item topic *Ti* to a certain domain this takes several attributes such as product category, company name, price and offers. This denoted the relevance of user T's personal interest to the category of item *i* in the RAS model by *RASu,i*

$$RASu,i = Sim(Tu,Ti).$$

This performs the similarity measure by fine filtered attributes. The derived products should satisfy the personal interest and as well as social influence without affecting their attribute consideration.

For each attribute in the attribute list (Ai) € sim (user1, user2, Ai)

From the above equation the similarity has been calculated between two users with the consideration of any particular attribute such as price, ie this measures whether the average price spent by user 1 is similar to user2, based on this the system suggests products to user 1 from user 2 log.

3) **Spammer Detection and Removal**

Review similarity is an important factor to determine spammer. Therefore, all detailed review behaviors of reviewer are calculated if the reviewer writes similar type of reviews. Rating spam score is also calculated for

that reviewer and the more spamming behaviors the reviewers make the more spamming scores they get. **Patten matching algorithm**

Input: Product Review

Output: pattern Pn

Matched string S

Steps:

1. Split the query Q into number of pattern P
2. Given a set of patterns, P1, P2, …, Pn ,
3. Give input document text T
4. find all occurrences of P in a text T = b1b2…bm.
5. do
6. if (text letter == pattern letter)
7. compare next letter of pattern to next
8. letter of text
9. else
10. move pattern down text by one letter
11. while (entire pattern found or end of text)
12. end

4. Dynamic Labeling using categorization

This research proposes categorization method to automatically assign each review into one of the three categories (i.e., early, majority and laggard). Given a product the review with an early on label is an early review and the user who wrote the review is an early reviewer. This starts with an examination of their post early reviews by look into average ratings of the reviews and concern scores voted by other user.

Optimal prescription finding process:

The best class label such as early, majority and laggard detection using categorization with Iteration algorithms

Input: D – E commerce data

T_N – Total Number

P – Pair generation attributes

C – Categories

S_c – contrast set group

Step 1: Read dataset from D

a) Read the attributes and values from T_N

b) Every attribute is set into 'P"

c) Set of id denoted as C

Step 2: Pre- process steps

Step 3: support and confidence calculation

a. identify base rules for every P from the following step

b. Sup (P) = (T(P))/n , Confidence (Pi➔Pi+1)

Step 4: find contrast set group Sc.

a. Set S_c.

b. If the property is already in the class label - find the probability

c. Else if new label, perform and assign in review

Step 5: check the threshold and probability values

Step 6: detect the threshold ratio from the dataset and return best class label such as early, majority and laggard.

5. DNE-BR (Predictive modeling)

Early reviews are certainly significant to product popularity. This modeling process completely avoid is a cold-start ranking problem. Proposed discriminate novel embedding-based ranking (DNE-BR) model which can effectively Predicting Early Reviewers. This Predictive modeling used to summarize the information of the entire text in a document.

DNE-BR Algorithm

Input training instances $\mathcal{T} = \{u \succ_p u' | u, u' \in \mathcal{U}\}$,
products embeddings set $\{v_p\}$,
learning rate λ,
margin coefficient m,
embedding dimensions L.
Output user embeddings $\{v_u | \forall u \in \mathcal{U}\}$

Procedure:
1: initialize user embeddings:
2: $v_u \leftarrow uniform(-\frac{6}{\sqrt{L}}, \frac{6}{\sqrt{L}}), \forall u \in \mathcal{U}$
3: $v_u \leftarrow v_u / ||v_u||, \forall u \in \mathcal{U}$
4: loop
5: sample a training instance $\langle u \succ_p u' \rangle \in \mathcal{T}$ do
6: update user embeddings:
7: $v_u := v_u - \frac{\partial \ell(\mathcal{T})}{\partial v_u}$,
8: $v_{u'} := v_{u'} - \frac{\partial \ell(\mathcal{T})}{\partial v_{u'}}$.
9: until convergence

This algorithm initially Rank the users simply based on the Number of Reviews (NR) that they have previously posted.
Rank the users based on the Number of times that a user has before acted as an Early Reviewer. Rank the entire users based on the Proportion that a user has acted as an Early Reviewer.Finaly it will return Early Reviewers list effective manner.

V. EXPERIMENTS AND RESULTS

5.1 DATA SETS

The key contributions of this paper are given as follows: this explores the use of objective features to model the subjective perception of n number of products review which has been received from every user. Specifically, a user-supplied training allows the system to determine Early Reviewers along with more efficiently the subjective Impact on Product Popularity.
This section describes the implementation process. Implementation is the realization of an application, or execution of plan, idea, model, design of a research. This section explains the software, datasets and modules which are used to develop the research.

5.1.1 SOFTWARE

The project experiments are performed on an Intel Dual Core with a RAM capacity 4 GB. The algorithms are implemented in ASP.Net for shopping site creation and C#.NET as coding language and are run under Windows family.The system has successfully implemented using Visual studio.Net and C#.net as code program.

5.1.2 Dataset:

The system mainly deals with the problem of cold start. The significant role of early reviews has paying attention in depth attention from selling practitioners to induce client purchase intention. So this need a set of data related to the shopping products with respective user profile and product shopping review. The data collected from real time websites such as Amazon, flipkart and snapdeal shopping, but some attributes are not available in the website, so the experiments taken the synthetic dataset which is described below.

productno	productcat	productname	productcost	productdes
1	Clothing and acc...	Wrist Watch	14999	Rich and Royal
10	Books	Ansi C	490	Complete knowledge in c
11	Books	Let us C	370	Complete knowledge in c
12	Books	Complete reference C++	625	Complete knowledge in c++
13	Clothing and acc...	Salwar kameez	890	Rich and Royal
14	Clothing and acc...	chudithar	650	Rich and Royal
15	Clothing and acc...	chudithar	750	Rich and Royal
16	Clothing and acc...	Salwar kameez	1500	Rich and Royal
17	Clothing and acc...	Saree	5980	Rich and Royal
18	Clothing and acc...	Saree	15720	Rich and Royal
19	Clothing and acc...	chudithar	600	Good look
2	Books	dotnet	300	dotnet black book
20	Clothing and acc...	Salwar kameez	750	Good look
21	Clothing and acc...	Salwar kameez	900	Good look
22	Clothing and acc...	handbags	599	Good look
23	Clothing and acc...	handbags	455	Good look
24	Clothing and acc...	handbags	550	Good look
25	Clothing and acc...	handbags	899	Good look
26	Clothing and acc...	handbags	980	Good look
27	Clothing and acc...	handbags	750	Good look
28	Clothing and acc...	handbags	699	Good look
29	Clothing and acc...	handbags	650	Good look
3	Books	java	250	java basics
30	Clothing and acc...	handbags	810	Good look
31	Clothing and acc...	handbags	585	Good look

Fig1: product dataset: product and its image

The above dataset shows the list of products available in the e-shopping website. This has been taken as a synthetic dataset for our current implementation. The implementation may consist n number of products for experiments. IRS has taken 500 products along with review details as initial dataset. As like the above dataset the system has another important dataset which creates more than 130 users shopping log.

After the synthetic data set is generated, and given the number of m local sites, each tuple from the synthetic uncertain database D is assigned to site Si chosen uniformly. Clearly, all local sites have the same data distribution. In particular, a local site server keeps a random sample set of the underlying data set, and the sample sets are mutually disjoint. In the experiments, every local server possesses an equal number of points, named the local cardinality.

5.2 Results and discussion

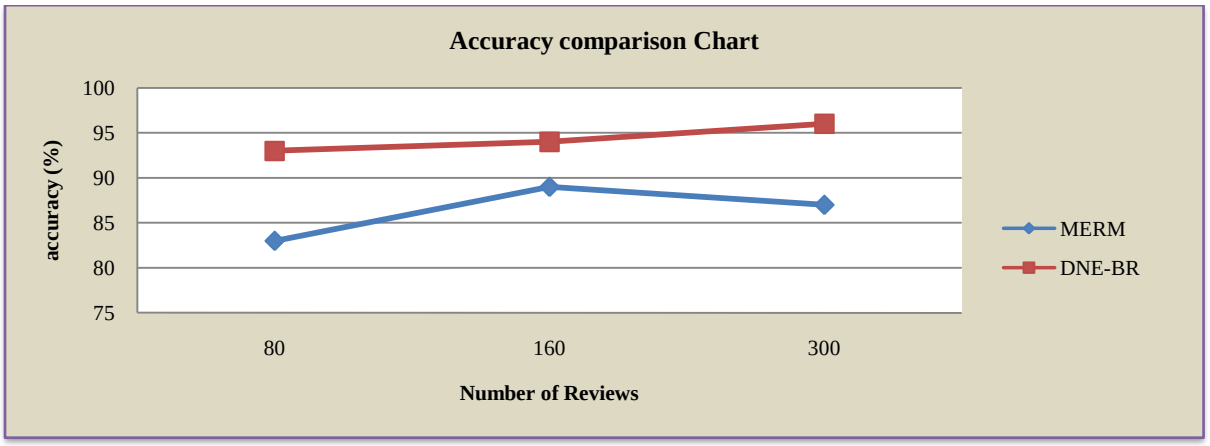
In each category of e shopping dataset, we use 80% of data as the training set and the remaining 20% as the test set. This is a synthetic dataset where the count can be increased by running the application every time. This chapter shows the results of every process in **DNE-BR** (Predictive modeling)**.**

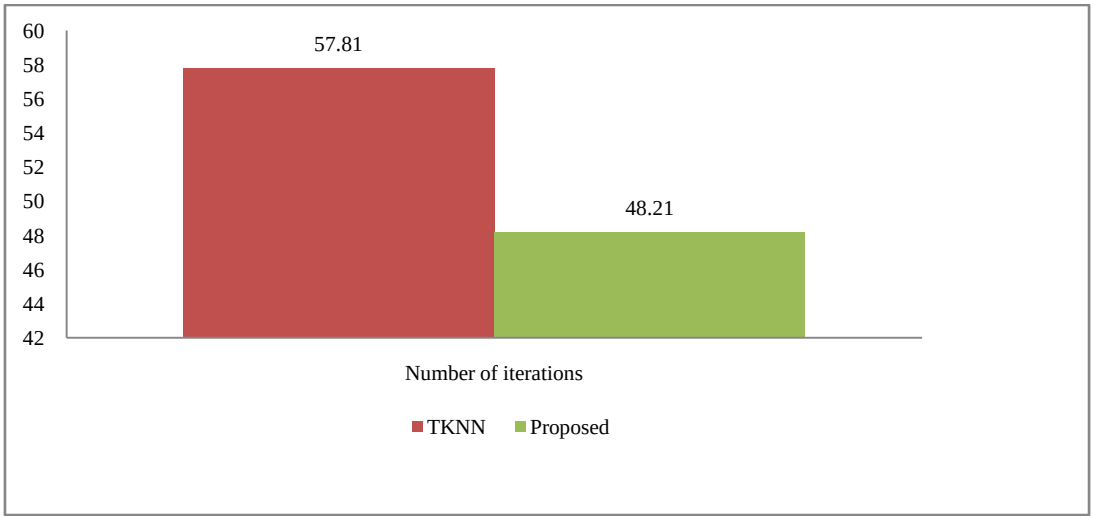
Phase 1: the phase1 starts with the profile analysis, where the user's basic information's has been collected. The followings are the registered user list in the e-shopping website.

5.3 Proposed DNE-BR Model with respect to Performance

The performance of this proposed work DNE-BR **(**Predicting Modeling**)** Scheme is compared with existing approaches MERM. The table shows the performance comparison of the proposed method with other existing approaches based on the four different metrics Accuracy, time, processing delay, number of iterations.

Metrics	Existing (MERM)	Proposed (DNE-BR)
Accuracy	90	98
Detection time	70.68	49.69
Delay	2984.06	2108.08
Number of iterations	57.81	30.21





From the chart it shows the performance measure based on the processing of the proposed approach took less number of iterations for the amount of datasets while comparing the existing method.

VI. CONCLUSION

The complete thesis provides an Early reviewer, Early review detection approaches on e-commerce datasets. In the literature, the authors were only concentrated on the Early review, the proposed system effectively finds the detect Early Reviewers along with Impact on Product Popularity with the help of DNE-BR. New approach named as DNE-BR has been proposed to find detect Early Reviewers accurately and fast. This system effectively apply buying behavior analysis model to identify and remove Spammer review .Experiments show this new DNE-BR have better results than MERM. The result is also useful for can eventually lead to the success of their new products.

6.1 Future Enhancement

In future work, the user early reviewer techniques will be extended with some other additional information. In the future, we will explore effective ways in

Incorporate review content information into our early reviewer prediction model. Also, this research work not concentrate studied the communication channel and social network structure in distribution of innovation partially due to the difficulty in obtaining the appropriate information from our review data. We will look into other sources of data such as ebay in which social networks can be extracted and carry out more insightful analysis. In this research remains an important issue to address how to improve product marketing with the identified early reviewers.

REFERENCES

[1]. Yoon-Joo Park, "Predicting the Helpfulness of Online Customer Reviews across Different Product Types", Seoul National University of Science and Technology, 2018.

[2]. SunilSaumyaa, Jyoti Prakash Singh, "Ranking online consumer reviews", Electronic Commerce Research and Applications, 2018.

[3]. Gerardo Ocampo Diaz and Vincent Ng, "Modeling and Prediction of Online Product Review Helpfulness: A Survey", Association for Computational Linguistics, 2018.

[4]. V. G. D. W. Shih-Lun Tseng, Shuya Lu, "The effect of herding behavior on online review voting participation," in AMCIS, 2017.

[5]. Subhabrata Mukherjee, Kashyap Popat, Gerhard Weikum, "Exploring Latent Semantic Factors to Find Useful Product Reviews, Society for Industrial and Applied Mathmatics, 2017.

[6]. Julian McAuley, Alex Yang, "Addressing Complex and Subjective Product-Related Queries with Customer Reviews", International World Wide Web Conference Committee, 2016.

[7]. Shih-Jung Wu, authorRui-Dong Chiang, Zheng-Hong Ji, "Development of a Chinese opinion-mining system for application to Internet online forums", The Journal of Supercomputing, 2016.

[8]. M. R. Islam and Minhaz F. Zibran, "Exploration and Exploitation of Developers' Sentimental Variations in Software Engineering", International Journal of Software Innovation, 2016.

[9]. W.D. J. Salganik M J, Dodds P S, "Experimental study of inequality and unpredictability in an artificial cultural market," ASONAM, 2016.

[10]. Arkaitz Zubiaga, Maria Liakata, Rob Procter, Geraldine Wong Sak Hoi, and Peter Tolmie, "Analysing how people orient to and spread rumours in social media by looking at conversational threads". PloS one, 2016.

[11]. Zhouyang Li, Lianzhong Liu, Chunfang Li, "Analysis of customer satisfaction from Chinese reviews using opinion mining", International Conference on Software Engineering and Service Science, 2015.

[12]. S.Atia and K. Shaalan, "Increasing the accuracy of opinion mining in Arabic", International conference on Arabic computing linguistics, 2015.

[13]. H.S. Le, T.V. Le and T.V. Pham, "Aspect Analysis for Opinion Mining of Vietnamese Text" International Conference on Advance Computing and Application, 2015.

[14]. T. Chumwatana, "Using sentiment analysis technique for analyzing Thai customer satisfaction from social media", International Conference on Computing and Informatics, 2015.

[15]. S.Ahmed and A.Danti, "A novel Approach for sentimental analysis and opinion mining based on sentiwordnet using web data", International Conference on Trends in Automation, Communications and Computing Technology, 2015.

Arifa.P." Analysis and Prediction of Early Reviewers and Product Popularity Using Discriminate Novel Embedding-Based Ranking (DNE-BR)." IOSR Journal of Engineering (IOSRJEN), vol. 09, no. 11, 2019, pp. 10-18