



IJIRCCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 11, Issue 5, May 2023

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.379



9940 572 462



6381 907 438



ijircce@gmail.com



www.ijircce.com

Real Time Online Products Review Summarizer

Vaidehi Sawant, Swaraj Pawar, Satvik Choudhari, Saket Deshmukh, Prof. Shashank Bholane

UG Student, Dept. of Computer Engineering, SCoE, Savitribai Phule Pune University, Pune, Maharashtra, India

UG Student, Dept. of Computer Engineering, SCoE, Savitribai Phule Pune University, Pune, Maharashtra, India

UG Student, Dept. of Computer Engineering, SCoE, Savitribai Phule Pune University, Pune, Maharashtra, India

UG Student, Dept. of Computer Engineering, SCoE, Savitribai Phule Pune University, Pune, Maharashtra, India

Asst Professor, Dept. of Computer Engineering, SCoE, Savitribai Phule Pune University, Pune, Maharashtra, India

ABSTRACT: Shopping was confined just to outdoor shopping a few years ago when there were no websites for online shopping and no internet. But now, the internet is available to everyone at fingertips with the advent of smartphones, tablets, laptops and even the cheaper rate to afford internet. This was the prime reason for the sudden booming of online shopping websites. With the coming of open business sectors and web based showcasing, there has been a gigantic measure of development in the online business area prompting extraordinary commitment both in monetary and non-monetary issues between the customers and organizations on the web. To improve the straightforwardness and client devotion alongside shopping experience, a large portion of the web-based retailers, for example, Amazon urges buyers to share their encounters, conclusions on the items or administrations that have been bought online. As the clients are allowed to offer their viewpoints, these audits tend to be extensive and have a couple of sentences that convey critical data about an item. Plus, the extended and typically pretentious sentences make it challenging for a planned purchaser to peruse and comprehend the intricate language utilized by the clients. What's more, only sometimes helps the likely purchaser in choosing whether to purchase a specific product or not. Text Rundown is a Characteristic Language Handling (NLP) strategy that concentrates and gathers information from the source and sums it up. Text summarization has turned into a prerequisite for some applications since physically summing up immense measures of data is troublesome, particularly with the growing greatness of information. Monetary exploration, site improvement, media observing, question answering bots, and report examination all benefit from text rundown. In this paper we are going to scrap the reviews first by beautiful soup and then use natural language tools to summarize the review. After extraction of reviews we are going to summarize the review. The extracted review, we are classifying the polarity like positive or negative by using random forest model

KEYWORDS: Natural Language Processing, Text Summarization, Customer Review, E-Commerce, Random Forest, Machine learning, TFIDF.

I. INTRODUCTION

Web based shopping has turned into something nowadays. Item surveys, rating of the item, fame of the item and nature of the item chooses what item to purchase from the entire part. An individual for the most part depends on item surveys and rating of the item for purchasing a specific item from the various kinds of the item shifting just concerning cost and quality having around something similar highlights. However, for a specific item, it is hard to go through every one of the surveys of the item. Simultaneously, I want to go through surveys and rating of all the results of the same kind and afterward need to reach a resolution. This frequently becomes an issue when the surveys are in enormous numbers. Innovation has definitely improved the part of human progress inside the last many years. In our way of life we are tending to parcels of text based item surveys information considered on the web, containing web destinations, news stories, announcements, and writings thus far. It will take the majority of our time during each day, time could be a fundamental thought about our life, so text synopsis might be a significant perspective to hack back time for perusing this text. Message summarization consequently creates a rundown containing significant sentences and incorporates all significant data from the first document. The objective of text synopsis is to accumulate unmistakable data from the source by sifting and giving a concise rundown. Until this point in time, a few procedures for text outline have been created. Text synopsis methods can be comprehensively arranged into four classifications: input, result, content and reason. Like from scrapping, summarization, Feature extraction and sentiment analysis. There are single and multi document rundown choices in light of the quantity of reports. There has been a consistent expansion in the quantity of

web clients consistently. With expansion in Web clients, comes a lot of data that gets put away online each second. This adds up to putting away gigantic measures of information each second. It might contain helpful and pointless information also. It requires countless server farms to store this tremendous measure of information. What's more, at times even the valuable information becomes hard to understand because of the commotion in them. Thus, there is a requirement for summing up this information without losing the first significance of the information and simultaneously decreasing the size of the information. Subsequently the course of Text Rundown comes into picture with its advantages spread over various fields, for example, AI, Regular Language Handling, Fake Learning, Semantics and so on, Not many years prior, when Web has not arrived at the everyday person, internet shopping was viewed as the most awful methods of shopping. Individuals never used to arrange online as it comes up short on contact and feel situations that we have when we go shopping. Product surveys have turned into a significant wellspring of information, not just for clients to track down suppositions about items and voice their remarks, yet in addition for makers to comprehend the criticism on their items. In advanced libraries, inventories have incorporated audit content both from arranged sources as well as from their supporters. Be that as it may, this abundance of data likewise makes it unwieldy; sense making on such an enormous assortment is troublesome, best case scenario on items with a huge number of audits. Manually fetching the reviews one by one is quite difficult. Faster decision-making. AI can make decisions and carry out actions far faster than humans, and latest generations of machine learning can consider facts and statistics as well as learn aspects of human emotion, then weigh both into its calculations.

A. Problem Definition:

We are going to summarize the lengthy reviews using NLP techniques and classifying the reviews through a random forest model.

B. Need:

This reduces the time required for grasping lengthy pieces such as articles without losing vital information.

II. RELATED WORK

Sainan Zhu, Jiaoju Ge et al. [1] stated that, this paper proposes a By utilizing numerous direct relapse model, they find the number of surveys, the score of items, the level of positive or negative audits, the quantity of expressions of audits, and the level of the observers in all actuality do have influences on the deals of item. Author additionally track down the internet based audit valence and volume of itself decidedly affect the business of its own items. The author used deep learning architecture such as artificial neural networks for model building.

Du Yanan, Yang Lu et al. [2] this paper proposes to break down the extremity and power of the fine- grained feeling in light of online audits of complicated items, and to check the contrast between the rating and the opinion of surveys, a feeling estimation model in light of a space philosophy, named DO-SSM (Sentiment Score Estimation model in light of the Area Metaphysics, DO-SSM) is proposed. The features like Review volume, Review valence, Number of people who find reviews helpful, Rating of the most helpful positive review, Qualitative characteristics of online review. The author used the neural network for model building.

Zhu YanChun,Zhang Wei [3] the paper makes a factual depiction for the information about business war stories, home video and the bar telephone on Amazon during from January to February of 2010. At long last, creator tracked down normal score of on the web audits might deceive the buyer buying choices and give the justification proposition to Web retailers and customers, while taking note of the commitments and deficiencies of this article, and examines the future exploration bearings.

Zhijie Zhao, Jiaying Wang et al. [4] this paper analyzes the factors and relative strength of online product sales based on four aspects: online reviews, review system curation, online promotional marketing, and seller guarantees. The empirical analysis of the SFNN model based on the data of Taobao.com shows whether the 14 factors, in relation to the four aspects, have any impact on product sales.

Ravali Boorugu, Dr. G. Ramesh et al. [5] stated that is an overview on the different kinds of text synopsis strategies beginning from the essential to the high level strategies. As per this review, the seq2seq model along with the LSTM and consideration component is utilized for expanded exactness.

III PROPOSED ALGORITHM

A. Design Considerations:

In a proposed system, we are proposing review summarization and classifying the summarized review by random forest model with a limited set of supervised data as shown in figure1. We are proposing Natural language processing tools to summarize the reviews. The NLP techniques for summarization for reviews. We are going to solve the accuracy issue in the diagnosis of summarization with accurate stage predictions. Information will be accumulated from different

spots, including web based business locales. The data assembled will be investigated and collected to create a corpus. Natural language processing (NLP) refers to the branch of computer science—and more specifically, the branch of artificial intelligence or AI concerned with giving computers the ability to understand text and spoken words in much the same way human beings can.

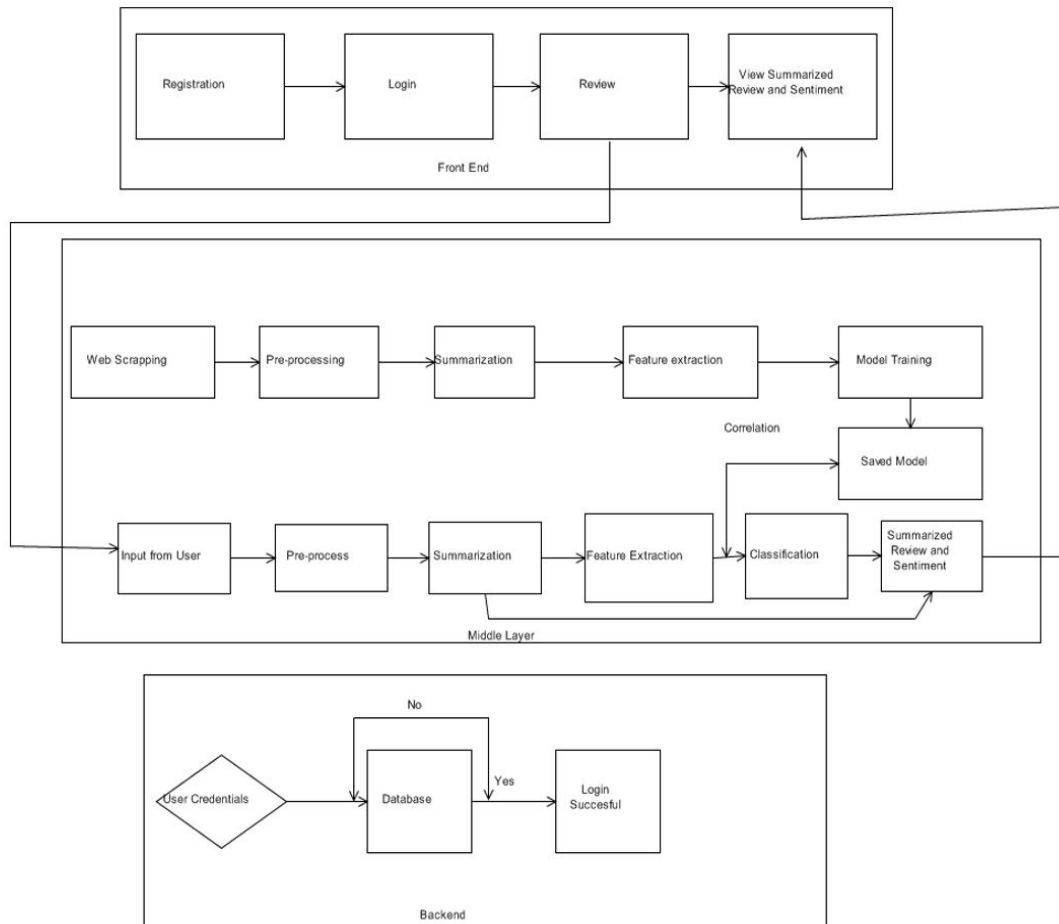


Fig. 1. Proposed Architecture

Step 2: *Selection Criteria:*

We are proposing Natural language processing tools to summarize the reviews. The NLP techniques for summarization for reviews. We are going to solve the accuracy issue in the diagnosis of summarization with accurate stage predictions. Information will be accumulated from different spots, including web based business locales. The data assembled will be investigated and collected to create a corpus. Natural language processing (NLP) refers to the branch of computer science and more specifically, the branch of artificial intelligence or AI concerned with giving computers the ability to understand text and spoken words in much the same way human beings can.

A. Beautifulsoup:

BeautifulSoup is a python package and as the name suggests, parses the unwanted data and helps to organize and format the messy web data by fixing bad HTML and present to us in easily-traversable XML structures. In short, BeautifulSoup is a python package which allows us to pull data out of HTML and XML documents.

B. Adding Corpus:

We are going to scrape the reviews from the ecommerce platform. After extraction we will store the data into csv/excel. The columns will be product_id, product, reviews and ratings.

C. Stop Words Removal:

After extraction of reviews from online platforms we will pre-process the data such as stop word removal. This was used to remove the unnecessary words and characters within each text, and creates a bag of words for the algorithms to compare against. The module 'Check Vectorizer' from Scikit-learn appoints numbers to each word/token while counting and gives its occasion inside a text. The case is summoned to forbid the English stop words, and these are the words, for example, A, In, The, Are, As, Is, etc, as they are not particularly helpful to order in text. This case is then fitted for the program to gain proficiency with the jargon.

D. Tokenization :

Tokenization is the method where the sentences within a text are broken into individual words (tokens). These tokens are saved into a cluster and utilized towards the testing information to distinguish the occasion of each word in a text. Stemming Stemming is the method involved with delivering morphological variations of a root/base word. Stemming programs are usually alluded to as stemming calculations or stemmers. Stemming is a significant piece of the pipelining system in Normal language handling. The contribution to the stemmer is tokenized words.

F. TF IDF:

TF-IDF stands for term frequency-inverse document frequency and it is a measure, used in the fields of information retrieval (IR) and machine learning, that can quantify the importance or relevance of string representations (words, phrases, lemmas, etc) in a document amongst a collection of documents (also known as a corpus). Term frequency works by looking at the frequency of a particular term you are concerned with relative to the document. Inverse document frequency looks at how common (or uncommon) a word is amongst the corpus. IDF is calculated as follows where t is the term (word) we are looking to measure the commonness of and N is the number of documents (d) in the corpus (D).

$$TF(t, d) = \frac{\text{number of times } t \text{ appears in } d}{\text{total number of terms in } d}$$
$$IDF(t) = \log \frac{N}{1 + df}$$
$$TF - IDF(t, d) = TF(t, d) * IDF(t)$$

G. Random Forest:

The random forest will classify the review category. An irregular timberland is an AI procedure that is utilized to take care of relapse and order issues. It uses group realizing, which is a procedure that joins numerous classifiers to give answers for complex issues. An irregular woodland calculation comprises numerous choice trees. The 'backwoods' created by the irregular timberland calculation is prepared through packing or bootstrap totaling. Packing is a group meta-calculation that works on the exactness of AI calculations. The (arbitrary backwoods) calculation lays out the result in light of the expectations of the choice trees. It predicts by taking the normal or mean of the result from different trees. Expanding the quantity of trees builds the accuracy of the result.

IV RESULTS

In this paper we are scraping data from flipkart. Around 10 products data we have scrapped using a beautiful soup module. After gathering data we have pre-processed the text data like stop word removal, upper to lower case, tokenization. After preprocessing we have extracted the features using the tf idf algorithm. After extracting the features we are trained on the model using random forest. We achieved 97.11% accuracy using random forest.

	precision	recall	f1-score	support
Negative	0.50	0.50	0.50	2
Positive	0.99	0.99	0.99	122
accuracy			0.98	124
macro avg	0.75	0.75	0.75	124
weighted avg	0.98	0.98	0.98	124

Fig. 2. Precision Vs Recall

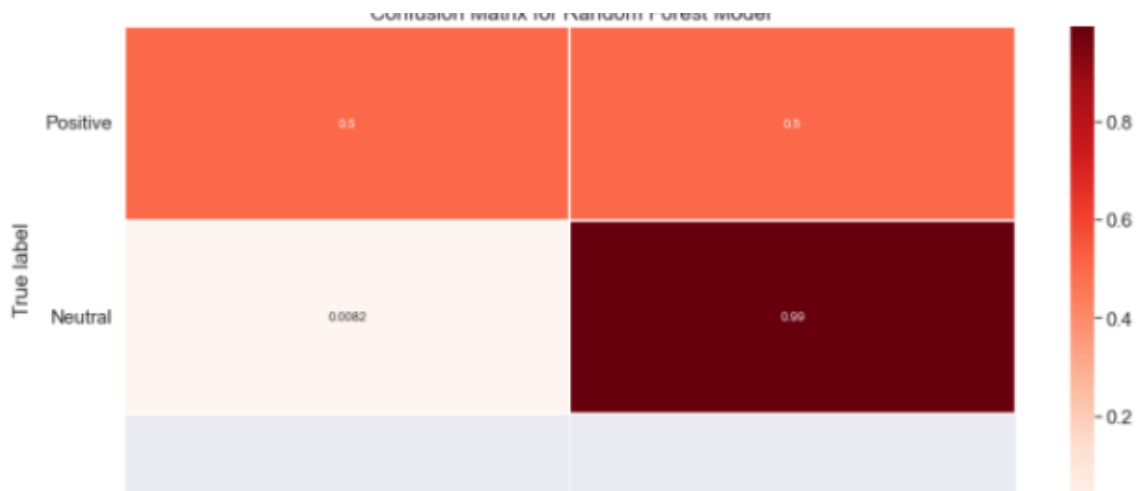


Fig. 3. Confusion Matrix

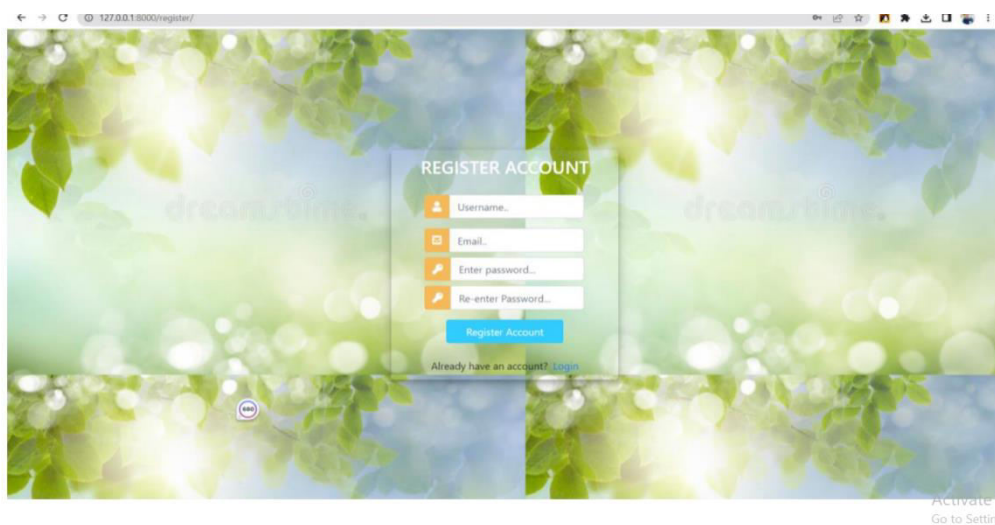


Fig. 4 Registration Window.

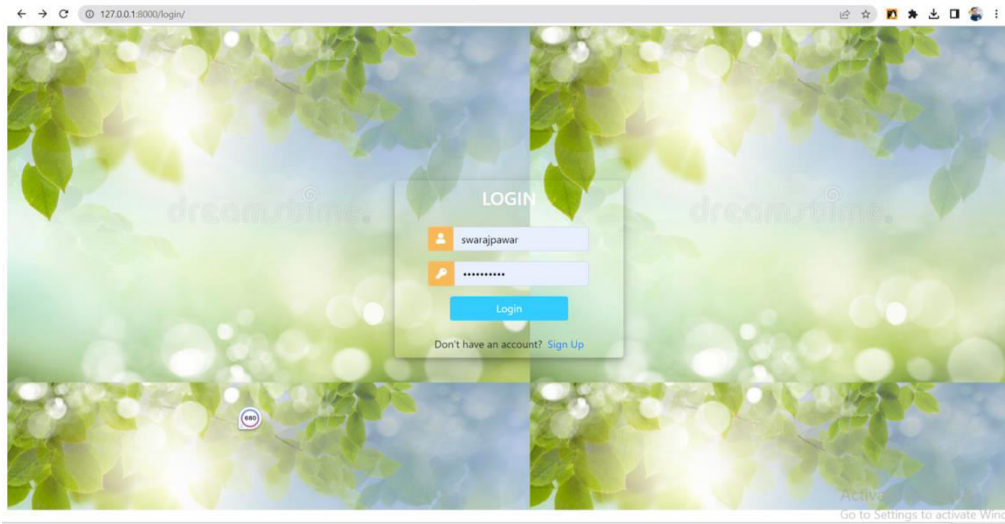


Fig. 5. Login Window

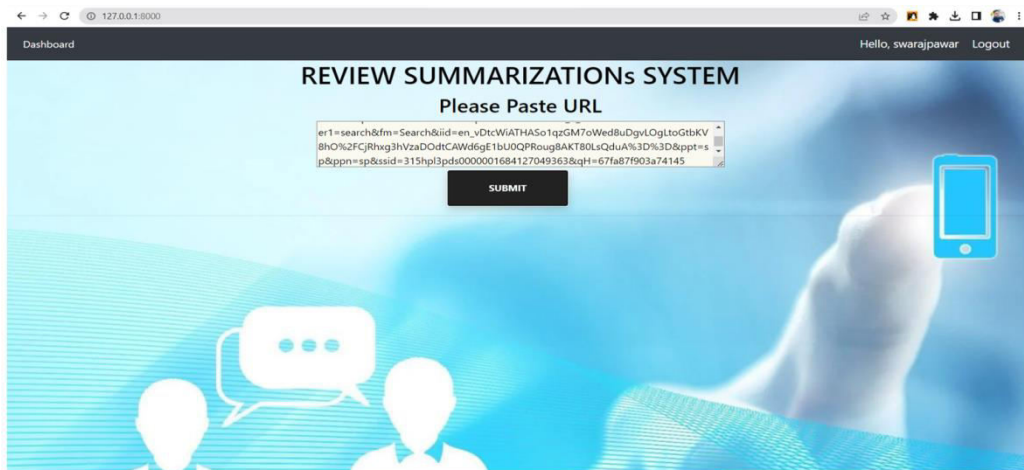


Fig 6. User input URL Window

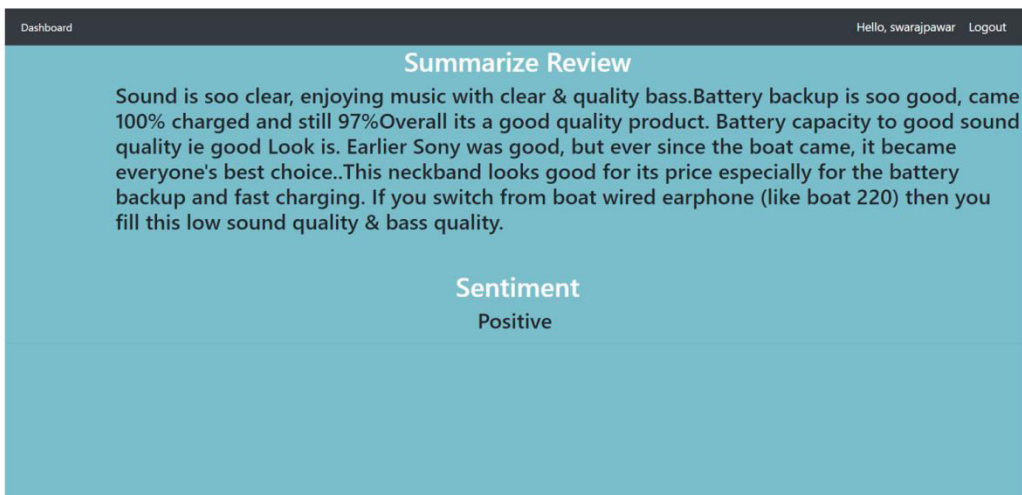


Fig 7. Output Window of summarized Review and its Sentiment

V CONCLUSION

The overview of the product based upon the user experience of a variety of different users. The summarized review that the users will get will be completely unbiased as the review will not depend on any one single user. After summarizing, the product review system will help to detect that review is positive or negative by random forest model.

REFERENCES

1. A. Y. L. Chong, E. Ch'ng, M. J. Liu, and B. Li, "Predicting consumer product demands via Big Data: the roles of online promotional marketing and online reviews," *International Journal of Production Research*, vol. 55, no. 17, pp. 5142- 5156, 2015.
2. F. Hou, B. Li, A. Y.-L. Chong, N. Yannopoulou, and M. J. Liu, "Understanding and predicting what influences online product sales? A neural network approach," *Production Planning Control*, vol. 28, no. 11-12, pp. 964- 975, 2017.
3. CNNIC "The 41st Internet Development Statistics Report," China Internet Network Information Centre. [Online]. Available: <http://www.cnnic.net.cn/hlwfzyj/hlwzbg/hlwtjbg/201803/t2018030570249.html>.
4. T. Tang, E. Fang, and F. Wang "Is Neutral Really Neutral? The Effects of Neutral User- Generated Content on Product Sales," *Journal of Marketing*, vol. 78, no. 4, pp. 41-58, 2014
5. K. Floyd, R. Freling, S. Alhoqail, H. Y. Cho, and T. Freling, "How Online Product Reviews Affect Retail Sales: A Meta- analysis," *Journal of Retailing*, vol. 90, no. 2, pp. 217- 232, 2014
6. W. Shi, X. Gong, Q. Zhang, and L. Wang "A Comparative Study on the First Time Online Reviews and Appended Online Reviews," *Journal of Management Science*, 2016.
7. Aditya Jain, Divij Bhatia, Manish K Thakur's —Extractive Text Summarization using Word Vector Embedding(2017).
8. Yang Wei and Yang Zhizhuo —Query based Summarization using topic background knowledge(2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)
9. Priya Pawar, Siddhesha Tandel, Shweta Bore & Nikita Patil —Online Product Review Summarization(2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIEC)
10. Paul Gigioli, Nikhita Sagar, Anand Rao, Joseph Voyles —Domain-Aware Abstractive Text Summarization for Medical Documents(published during 2018 IEEE BIBM.
11.] Shivangi Modi & Rachana Oza —Review on Abstractive Text Summarization Techniques (ATST) for single and multi documents(2018 International Conference on Computing, Power and Communication Technologies.



Impact Factor: 8.379



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

 9940 572 462  6381 907 438  ijircce@gmail.com



www.ijircce.com

Scan to save the contact details