iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** XII **Month of publication:** December 2021

DOI: https://doi.org/10.22214/ijraset.2021.39382

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue XII Dec 2021- Available at www.ijraset.com

Deep Learning Approach of Product Evaluation Using Comment Analysis

Syed Mudasar[1], Dr. Jasmeen Gill[2]
[1, 2]*RIMT University, Mandi Gobindgarh Punjab*

Abstract: ***Digital reviews now play a critical role in strengthening global consumer communications and influencing consumer purchasing patterns. Consumers can use e-commerce giants like Amazon, Flipchart, Snap deal, Jio and others to share their experiences and provide real insights about the performance of a product to future buyers. The classification of reviews into positive and negative sentiment is required in order to derive relevant insights from a big set of reviews. Comment Analysis is a computer programme that extracts subjective data from text. Out of Various Classification models Deep Learning Approach of Product Evaluation Using Comment Analysis is to develop a model that uses AI technologies like Deep Learning to process thousands and millions of online reviews on a product in a split second of time and rate the products on a scale of 1-5 based on the user comments We have worked on two deep learning models based on Recurrent Neural Networks (RNN) and Graph Convolution Network (GCN).***

Keywords: LSTN, GCN, NLTK.

I. INTRODUCTION

Everyone nowadays is aware with the concept of purchasing or using internet products. And, as the demand for "things" grows, it becomes increasingly important for the buyer to understand the pool into which he is wading. Is it a source of genuine content, or is he about to become the latest victim of deceptive advertising?

With such a large number of e-commerce websites up and operating on the Internet, and with it, an even larger number of things available online, determining which is the true worthy-to-be-bought commodity has become nearly impossible Current market conditions make it quite easy to develop any type of product and sell it online. Now how will the consumers know about the quality of the product, the durability, and all other relevant factors.

To answer this question, technologists all across the world devised the brilliant concept of "product rating" Everything and anything may be simply rated. You purchase a product and then rate it on the website where you purchased it. When you take a taxi, you have the option of rating both the taxi service and the driver [1]. The technological advancements did not end there. Techies have now devised the concept of feedback and suggestions. You can now not only rate your product on a 5-scale bar, but can also give any feedback or complaints of your choice. But now there's a problem with the feedback system. On a single website, there are thousands and millions of feedbacks and reviews on a single product [2]. How does one go about reading all of those reviews and then processing them in order to arrive to a conclusion? The machine, on the other hand, can read all of those reviews in less than a second and even process them to come up with a conclusion. In this research we use the same concept to solve the what-to-buy problem even further. The Deep Learning Technique is used to solve this problem in this research “Product Evaluation using comment analysis” provided here [3]. The algorithm created here analyses all of the product reviews on a website, processes them, and gives the customer a fresh 5-star rating bar. Overall, users now have access to two sets of rating scales: One that is rated by other users, and the other generated by this model that is based on the reviews provided on that product.

The organization of paper is as follows: a. explicates the approach followed to conduct analysis b. demonstrates [4] the dataset used .c Shows the data preprocessing technique that is Data Extraction and Data Cleaning. (d) Explains the Machine Learning Model i.e., Recurrent Neural Networks (RNN) Model and Graph Convolution Network (GCN) Model and their experimental results of the models and cross validation of predictive accuracy of models. Finally, section (E) concludes the proposed work and describes its future scope. The main objective of our project is to harness the powers of AI technologies like Machine Learning and Deep Learning to develop a model that processes thousands and millions of online reviews on a product in a split second of time. More precisely, the Recurrent Neural Network and Graph Convolution [5] Network technique have been used in this project to achieve the goal. In Recurrent Neural Networks (RNN), we have used Long Short-Term Memory (LSTM) Model [6] to capture the sequential behavior of comments. In Graph Convolution Network (GCN),[7] we build a large and heterogeneous comment word graph which contain word nodes and comment nodes so that global word co-occurrence can be explicitly modeled and graph convolution can be easily adapted. The model would generate a rating out of 5, where 1 means the product is devoid of quality, and not worth buying, while 5 means the quality is up to the mark and is worth buying to the consumers.

©IJRASET: All Rights are Reserved

The model first reads all the reviews and gives an internal rating to each review [5].

After all the reviews are done, the model now finds the average of all the internal ratings predicted by it. This is the final Output that we present. As far as the training dataset is concerned, the dataset was fetched from the "Amazon Product Data" [4].

The comments were then used by our model to classify the product between 1 star and 5 stars. The dataset consists of 142.8 M comments, out of which the model has been trained and tested for a total of 125,000 comments. The dataset will be divided in a 70-30 fashion for training and testing purposes.

The organization of paper is as follows: Section 3 explicates the approach followed to conduct analysis (Problem Statement) and the Objective. Section 4 demonstrates the dataset used and shows the data preprocessing technique that is Data Extraction and Data Cleaning. And explains the Machine Learning Model i.e., Recurrent

Neural Networks (RNN) Model and Graph Convolution Network (GCN) Model and their experimental results of the models and cross validation of predictive accuracy of models. Finally concludes the proposed work and describes its future scope.

1) *Dataset:* The dataset being used in this project is the "Amazon Product Data" distributed by Julian McCauley from University of California [6]. This dataset contains product reviews (ratings, text, helpfulness votes) and metadata (Descriptions, category information, and price, brand, and image features) from Amazon, including 142.8 million reviews [3]. We have used five categories for Model training and testing i.e., Electronics, Fashion, Appliances, Clothing and Home.
2) *Data Preprocessing:* Data preprocessing is the technique of the machine learning (cleaning and organizing) in which the raw data is converted into suitable and structured data [4] which will be suitable for building and training Machine Learning models.
3) *Data Extraction:* We did extract only those comments [3] in which both the field's i.e., 'review Text' and 'overall 'are present. So, the final data contains equal no of comment among all classes
4) *Data Cleaning:* Remove stop words: Used nltk (Natural language tool kit), python library to download stop words. Iterated all the comments from the dataset and removed those words [2] which are stop words. Creating vocabulary: Iterated all the comments from the dataset again and removed those words that repeated in data less than a threshold number, i.e., 5 and Removed symbols and special character

II. LITERATURE REVIEW

Tanjim Ul Haque, et.al (2018), studied final comparative analysis was done on this research paper, it was found that the proposed model gained a 93.5% accuracy for cellphones, electronics, and their respective accuracies. For musical instruments, the proposed model bagged a high accuracy of 94.02% [9]. It was found that SVM provided the best classifying results in this case. Disadvantage: This model gives only two outputs. That is positive and negative whereas our model gives a rating on a scale from 1-5. Hence the consumers can make a better decision using our model.

Shadi AlZu'bi, et.al (2019), classified the comments as positive and negative. Natural Language Processing was used for the task [10]. The team focused on the helpfulness/unhelpfulness of the comment.

Zeenia Singla, et.al (2018) conducting sentiment analysis of mobile phone reviews they classified the comments as positive [11] and negative comments the team focused only on the negative and positive comment and provide the out of the them, they have not worked on rating bar.

Soonh Taj, et.al (2019) considered that there are numerous headings in feeling examination that can be investigated. This paper investigated opinion examination of news and web journals utilizing a dataset from BBC involving new articles. It was seen that classes of business and sports had increasingly positive articles, while amusement and tech had a lion's share of negative articles. Future work right now be founded on slant investigation of news utilizing different AI approaches with the advancement of an online application from where clients can peruse updates on their inclinations. Additionally, in light of assessment investigation strategies, pursuers can tweak their news channel [12].

Wataru Souma, et.al (2019) proposed the prescient intensity of verifiable news assessments dependent on monetary market execution to gauge money related news slants. We characterize news assessments dependent on stock value returns found the middle value of more than brief just after a news story has been discharged. In the event that the stock value shows positive (negative) return, we order the news story discharged only before the watched stock return as positive (negative). We use Wikipedia and Gig word five corpus articles from 2014 and we apply the worldwide vectors for word portrayal technique to this corpus to make word vectors to use as contributions to the profound learning TensorFlow system. We find that the estimating exactness of our technique improves when we change from arbitrary determination of positive and negative news to choosing the news with most elevated positive scores as positive news and news with most noteworthy negative scores as negative news to make our preparation informational collection [13].

©IJRASET: All Rights are Reserved

A Shalkarbayuli,et.al (2018) aimed that Traditional strategies for content arrangement works normal on Russian messages and utilizing TF-IDF works genuinely well on slant investigation. Despite the fact that it requires a ton of pre-processing and model structure it don't have constraints that have in most Google Services. Conventional calculations like SVM and Naive Bayes works generally well on content grouping if TF-IDF is utilized as an inserting.

As detailed over these 2 calculations are inverse of one another. SVM with high exactness on negative and positive classes while, Naive Bayes has high accuracy on nonpartisan class [14].

Mika V. Mantyla,et.al(2018) aimed that the historical backdrop of feeling examination, assessed the effect of opinion examination and its patterns through a reference and bibliometric study, delimited the networks of notion investigation by finding the most mainstream distribution setting, found which look into themes have been researched in conclusion investigation, and checked on the most referred to unique works and writing audits in slant investigation. In this manner, notion investigation is likewise having an effect in any event when estimated by the quantity of citations. Sentiment examination had utilized numerous information sources identified with or originating from papers, tweets, photographs, visits for instance [15].

Vaanchitha Kalyanaraman ,et.al (2017) investigated the causative connection between news stories and estimation of the cost of stocks in the market. Our AI model had the option to foresee the supposition of an article with exactness of 53.2% utilizing Normal Equation and 59.5% utilizing Gradient Descent when contrasted with the outcome physically anticipated by us. On examination with real stock costs likewise, we found that Gradient Decent was increasingly exact with an exactness of 81.82% while Normal Equation had a precision of just 54.54%. In the two cases, we plainly observe Linear Regression utilizing Gradient Descent to be increasingly productive [16].

Kia Dashtipour, et.al (2016) studied an outline of cutting-edge multilingual assumption investigation strategies. It depicted information pre-handling, ordinary highlights, and the principal assets utilized for multilingual slant examination. At that point, examined various methodologies applied by their creators to English and different dialects. We have arranged these methodologies into corpus-based, vocabulary based, and cross breed ones [17].

Shuhaida Mohamed Shuhidan, et.al (2018) study covers the execution of AI calculation approaches in slant examination of Malaysia monetary news features. This investigation can be utilized for partners who need to think about the money related news and look for information or information in the monetary world.

The information are p e p o l s o l e o e el e e s e o s ess e o e s es e o ppl es p o e o - se l l o e Bayes calculation as the strategy to perform feeling investigation. This examination comprises of a few stages in pre-handling, for example, separate information, stop word expulsion, and stemming to clean the dataset and make it as information planning before playing out the assumption investigation with the chose AI calculations. In the stop word expulsion, bundle in R is utilized to clean the dataset while for stemming process, Snowball stemmer is utilized to set the information to its root word. Test results of investigation are clarified for the two calculations.

The end depicts the summation of the examination and future works [18]. 11.Ubale Swati, et.at (2015) Determining the disposition of an author with Respect to some subject or the general inclination in a report is fundamental point of doing slant examination. News examination can be utilized to plot the company's conduct after some time and accordingly yield significant key bits of knowledge about firms.

Assessment investigation is additionally helpful in online networking checking to consequently portray the general inclination or state of mind of shoppers as reflected in web-based life toward a particular brand or organization and decide if they are seen emphatically or adversely. Countless organizations use news investigation to assist them with settling on better business choices so in our task we are doing assumption examination on news story identified with organization [19].

III. RESEARCH METHODOLOGY

The process followed in my work consists of two steps, we have created two machine learning model which is bases of Recurrent Neural Networks [8] and Graph Convolution Network. In Recurrent Neural Networks (RNN), we have used Long Short-Term Memory (LSTM) Model in order to capture the sequential behavior of comments and in Graph Convolution Network (GCN), we build a large and heterogeneous comment word graph which contain word nodes and comment nodes so that global word co-occurrence can be explicitly modeled and graph convolution can be easily adapted.

A. Recurrent Neural Networks (RNN) Model

Recurrent Neural Networks are Deep Learning models with simple structures and a feedback mechanism built-in, or in different words, the output of a layer is added to the next input and fed back to the same layer.

©IJRASET: All Rights are Reserved

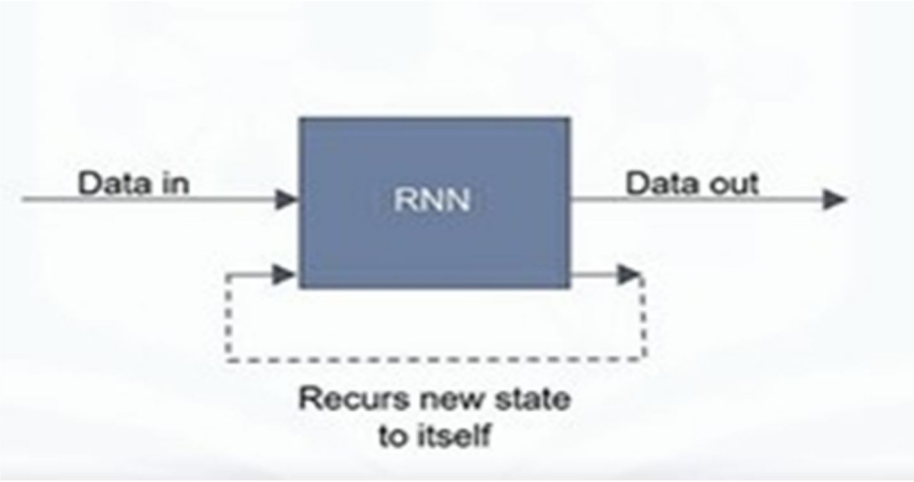


Fig 1: Recurrent Neural Networks Model

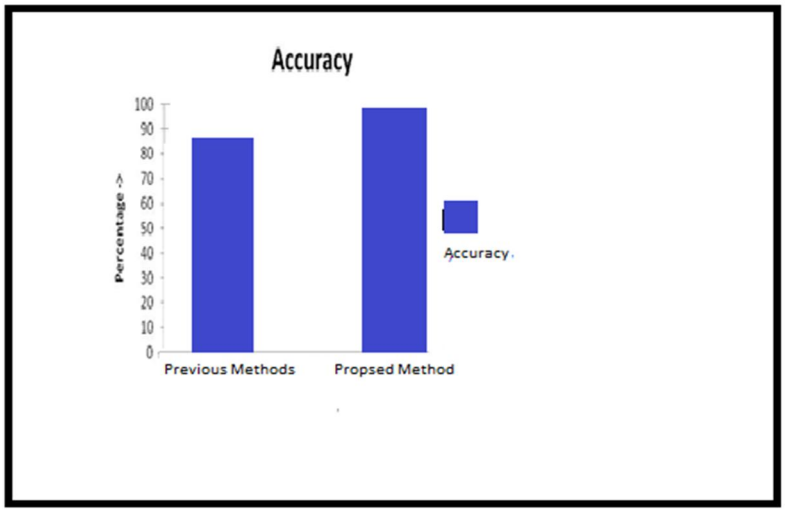


Fig 2: Accuracy

IV. CONCLUSION AND FUTURE SCOPE

A. Conclusion

Due to the evolutionary shift from offline markets to online market, the digital market has as increased the dependency of customers on online reviews to a great extent. Online reviews have become a great platform for building trust of the consumer buying patterns. Our research is aiming to achieve this by conducting PRODUCT EVALUATION USING COMMENT ANALYSIS Using the Machine learning model of Long Short-Term Memory (LSTM Model) and Graph Convolution Network (GCN) Model. *After* the Implementation the accuracy results have been cross validated and the highest values of accuracy [3] achieved for Long Short-Term Memory (LSTM Model) are Trained at learning rate 0.001 for 10 epochs. Achieved maximum training accuracy 98% percent. Achieved maximum validation accuracy 97.3% percent. Achieved minimum loss during training 0.932. Achieved minimum loss during validation 0.94. Achieved maximum test accuracy 62 percent and for the Graph *Convolution Network (GCN) Model* the highest values of accuracy achieved are; trained the model at learning rate 0.001 for 110 epochs. Achieved maximum training accuracy 68 percent. Achieved maximum validation accuracy 58 percent. Achieved minimum loss during training 0.82811. Achieved minimum loss during validation 1.0005. Achieved maximum test accuracy 58.5 percent

The research is helpful in Product evaluation using comment analysis which makes E-commerce platform a better place for shopping

©IJRASET: All Rights are Reserved

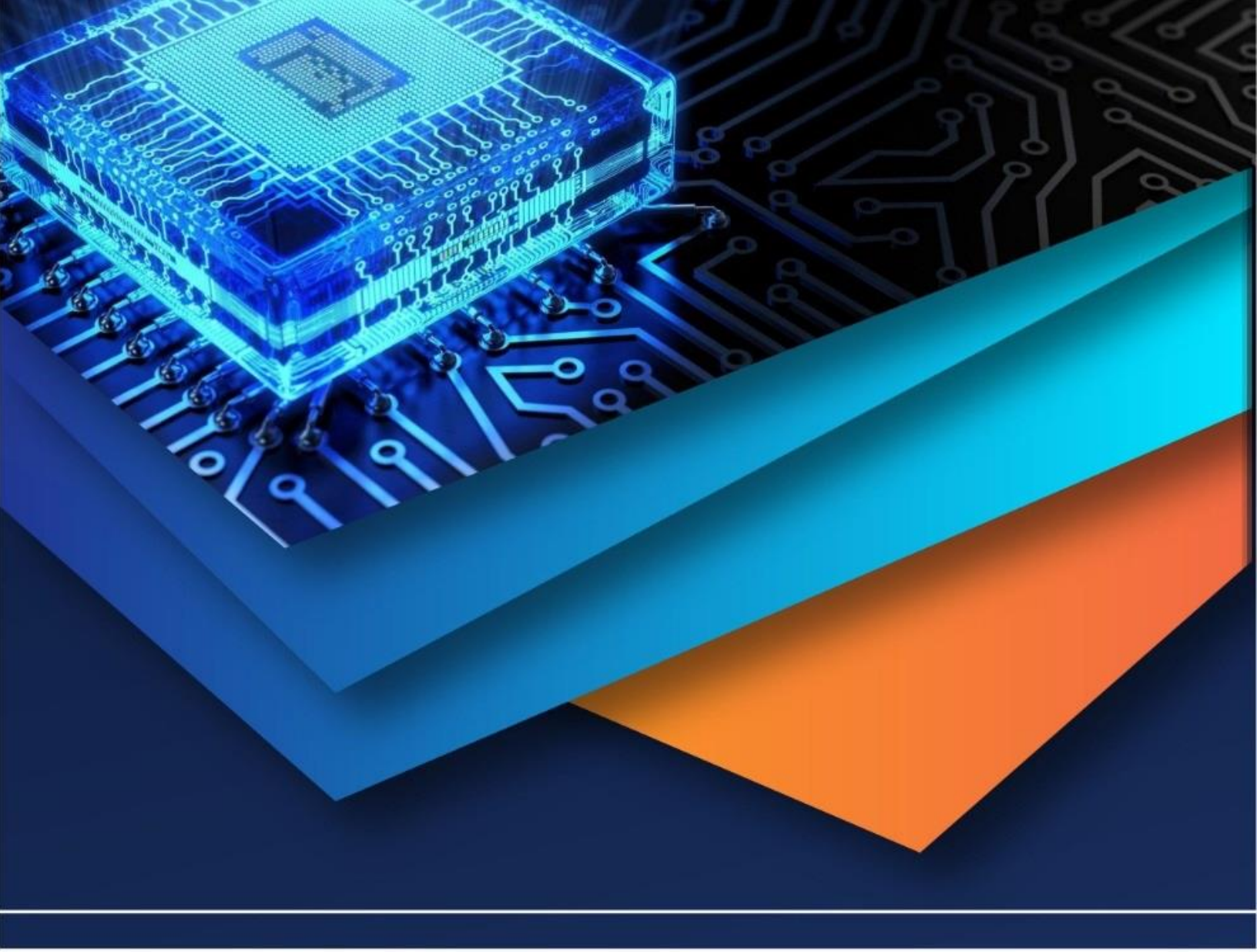
B. Future Scope

Algorithms related to Graph Convolution Network is one of the most recent algorithms in the Machine learning community, so there is a lot of space for research in this field over different applications. Combination of Recurrent
Neural Network and Graph Convolution into one model can improve the accuracy. Dataset is large enough and wide, there are many fields in the dataset which can be used in natural language project.

REFERENCES

[1] Cai, H.; Zheng, V. W.; and Chang, K. 2018. A comprehensive survey of graph embedding: problems, techniques and applications. IEEE Transactions on Knowledge and Data Engineering 30(9):1616– 1637.
[2] Frank Fuhlbruck , Johannes Kobler , Ilia Ponomarenko, Oleg Verbitsky.The Weisfeiler- Leman Algorithm and Recognition of Graph Properties.
[3] Thomas N. Kipf & Max Welling 2017. Semi-Supervised Classification with graph convolutional networks.
[4] Y. Kim, "Convolutional Neural Networks for Sentence Classification," Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP 2014), pp. 1746–1751, 2014.
[5] A. L. Maas, R. E. Daly, P. T. Pham, D. Huang, A. Y. Ng, and C. Potts, "Learning Word Vectors for Sentiment Analysis," Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (HLT 2011), pp. 142-150, 2011
[6] Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up? Sentiment classification using machine learning techniques. In: Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing-Volume 10. Association for Computational Linguistics, 2002 [14].Turney, P.D.: Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics. Association for Computational Linguistics, 2002
[7] Hu, M., Liu, B.: Mining and summarizing customer reviews. In: Proceedings of the tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2004
[8] Devika, M.D., Sunitha, Cª., Ganesh, A.: Sentiment analysis: a comparative study on different approaches. Proc. Comput. Sci. 87, 44–49 (2016)
[9] Bhadane, C., Dalal, H., Doshi, H.: Sentiment analysis: measuring opinions. Proc. Comput. Sci. 45, 808–814 (2015)
[10] Rosenthal, S., Farra, N., Nakov, P.: SemEval-2017 task 4: sentiment analysis in Twitter. In:
[11] Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017) [20]. Fang, X., Zhan, J.: Sentiment analysis using product review data. J. Big Data 2(1), 5 (2015) Ahmad, M., et al.: Hybrid tools and techniques for sentiment analysis: a review. Int. J. Multidiscip. Sci. Eng. 8(3) (2017)
[12] Soonh taj, Baby Bakhtawer shaikh, "Sentiment analysis of News Articles(A Lexicon based approach)", 2019,IEEE.
[13] Wataru souma, Irena Vodenska, Hideaki Aoyana "Sentiment "analysis using deep learning methods", 2019,Springer.
[14] A shalkarbayuli "Sentiment analysis review", 2019, IEEE.
[15] Mika V. Mantyla, Daniel Graziotin,Mikha Kuratila "The evolution of sentiment analysis" ,2018, Elsevier.
[16] Vaanchitha Kalyanaraman, Sarah Kazi, Rohan Todulkar,Sangeeta oswal "Sentiment Analysis on News Articles for stocks",2014, IEEE.
[17] Kia Dashtipour,Mandar Gogate,Ahsan adeel,Cosimo Leracitano, Hadi lacijani, Amir Hussain "Exploiting Deep Learnining for Persian sentiment analysis",2018 ,Springer.
[18] Hhaida Mohamed Shuhidan, Saidatul Rahah Hamidi, Soheil Kazemian, Shamila Mohamed Shuhidan, Maizatul Akmar Ismail "Sentiment Analysis for Financial News Headlines using Machine Learning Algorithm",2018 ,Springer.
[19] Swati Redhu, Sangeet Srivastava, Barkha Bansal, Gaurav Gupta, "Sentiment Analysis Using Text Mining: A Review",2018, International journal of data science and technology

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)