iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** VII **Month of publication:** July 2021

DOI: https://doi.org/10.22214/ijraset.2021.36202

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 VII July 2021- Available at www.ijraset.com

Sentiment Analysis of Customer Product Reviews using deep Learning and Compare with other Machine Learning Techniques

Amit Purohit[1], P. S. Patheja[2]
[1]*Department of School of Computer Science, Research Scholar, VIT, Bhopal*
[2]*Department of School of Computer Science, Professor VIT, Bhopal*

Abstract: ***Sentiment analysis is defined as the process of mining of data, view, review or sentence to Predict the emotion of the sentence through natural language processing (NLP) or Machine Learning Techniques. The sentiment analysis involve classification of text into three phase "Positive", "Negative" or "Neutral". The process of finding user Opinion about the topic or Product or problem is called as opinion mining. Analyzing the emotions from the extracted Opinions are defined as Sentiment Analysis. The goal of opinion mining and Sentiment Analysis is to make computer able to recognize and express emotion. Using social media, E-commerce website, movies reviews such as Face book, twitter, Amazon, Flipkart etc. user share their views, feelings in a convenient way. Sentiment analysis in a machine learning approach in which machines classify and analyze the human's sentiments, emotions, opinions etc. about the products. Out of the various classification models, Naïve Bayes, Support Vector Machine (SVM) and Decision Tree are used maximum times for the product analysis. The proposed approach will do better result as compare to other machine learning techniques.***

Keywords: Opinion Mining, Sentiment Analysis, Naïve Bayes classifier, Logistic Regression, Svm, Knn, Product Review

I. INTRODUCTION

The main goal of research related to sentimental analysis is to obtain author feeling expressed in positive or negative comment. In this paper, a detailed survey of different techniques or approach is used in sentiment analysis to gain useful insight into the data even when you have access to it. This analysis is performed in multilevel i.e Document level, Sentence level and aspect level. This topic is very challenging –although although a lot of work has been done in this field, accuracy is still rather average due to comment, slang, smile etc.Social media plays very important role in almost everybody's day to day life. It allows the people to convey what they think and feel about the products in E-commerce determine the mood of the writer or attitude of the speaker; it may be either positive or negative towards the product. In [1], Jayanti Desai these positive or negative emotions expressed by the people are known as sentiment. Opinion mining or sentiment analysis refers to the type of Natural Language Processing (NLP), Text analysis and Computational-Linguistics to identify and extract subjective information in source material. In 4], E.A.Vimal4 since the online data's are tremendously growing data-by-day, it is considered to be very important in the current situation because, lots of user opinionated texts are available in the web now. Sentiment analysis is considered to be the study of user's thought and feeling towards a product. Both SA and OM are interchangeable. The importance of the sentiment analysis or opinion mining is increasing day by day, as data grows day by day. Machines must be reliable and efficient to interpret and understand human emotions and feelings. Since customers express their thoughts and feelings more openly than ever before, sentiment analysis is becoming an essential tool to monitor and understand that sentiment. In [2], focuses on review mining and sentiment analysis on Amazon website. Users of the online shopping site Amazon are encouraged to post reviews of the products that they purchase. Amazon employs a 1-to-5 scale for all products, regardless of their category, and it becomes challenging to determine the advantages and disadvantages to different parts of a product. Automatically analyzing customer feedback, such as opinions in survey responses and social media conversations, allows brands to learn what makes customers happy or frustrated, so that they can tailor products and services to meet their customers' needs. As with any method, there are different ways to train machine learning algorithms, each with their own advantages and disadvantages. To understand the pros and cons of each type of machine learning, we must first look at what kind of data they ingest. In ML, there are two kinds of data — labeled data and unlabeled data. Labeled data has both the input and output parameters in a completely machine-readable pattern, but requires a lot of human labor to label the data, to begin with.

Unlabeled data only has one or none of the parameters in a machine-readable form. In [9], the supervised learning approach and Dictionary based techniques are considered for the Sentiment analysis. The precision and recall measures are used to evaluate the accuracy of the algorithm. The survey paper in [10] discuss about the overview of different classification, clustering algorithms and challenges in sentimental analysis and opinion mining. This negates the need for human labor but requires more complex solutions. The reviews have been classified using machine learning classification models like Naïve Bayes, Support Vector Machine (SVM) ,Decision Tree and and compare with CRNN model which used for product customer review using sentimental analysis. In Recent Years, the Deep Learning Method Based on Artificial Neural Network Has Developed Rapidly, Which Provides A New Idea for Emotion Classification

II. RELATED WORK

The contextual polarity of the phrases was taken into consideration and ambiguity was removed [5].Also, a refined method has been devised to establish contextual polarity of phrases by using subjective detection that compressed reviews while still maintaining the intended polarity [6]. Delineated study has been conducted on tweets available on Twitter, movie reviews to build the grounds on sentiment analysis and opinion mining. A sentiment classifier has been built to categorize positive, negative and neutral sentiments not only in English but also for other languages using corpus from Twitter [7]. The polarity of Smartphone product reviews has been found only on the basis of positive and negative orientation of the review [8]. A system has been built using support vector machine where sentiment analysis is carried out by taking into consideration sarcasm, grammatical errors and spam detection [9]. An enhanced Naïve Bayes model by combining methods like effective negation handling, word anagrams and feature selection has been utilized to conduct sentiment analysis [10]. Sentiment analysis is not only confined to the English language but has been implemented for various languages. Sentiment analysis of Chinese text by implementing four feature selection methods and five classifiers viz. Centroid classifier, K-nearest neighbor, Window classifier, Naïve Bayes and SVM has been done [11]. One of the solutions to Sentiment Analysis namely Opinion Digger was introduced by Moghaddam and Ester [12]. This unsupervised Machine Learning methodology works at Sentence level. Correlates and compares product aspect and standard rating guidelines (used in Amazon, Snapdeal, flipkart1 etc). This proposed work is divided into two sub methods. At first, input information is fragmented into sentences. Repeated nouns in the sentences are coined as aspects. Aspect (repeated nouns) if forms any pattern, are stored. Secondly, aspects are compared to the rating guideline (like 4 means “Good”, 3 means “Average”, etc) and accordingly labeled as “Good” ,”Average” and ”Bad” [12]. Major advantage is its high performance in product rating at aspect level with a loss of 0.49 only. A Joint Model of Feature Mining and Sentiment Analysis for Product Review Rating was presented by de Albornoz et al. [13]. This machine learning method rates product at global level considering whole opinion at once. This approach is basically carried out in four steps. Sentiment Analysis is not limited just to reviews or Twitter data but is also applicable on stock markets [14]), news articles [15] or political debates [16]. Sentiment analysis can be used to flourish consumer products related business [17]. It uses rule-based approach for sentiment analysis to extract topic words of negative opinion sentences and thus promote the competitors of the products receiving negative feedback. Similarly, relevant ads based on a person’s liking or disliking are displayed on various blogging sites targeting bloggers. Blogger-Centric Contextual Advertising Framework has been concocted to determine users’ personal interests and display those ads that intersect with them [18]. Deep Learning uses powerful neural network algorithms to mimics the way human brain process data for translating languages, recognizing speech, detecting objects and making decisions. Deep Learning algorithms are able to identify and learn the patterns from both unstructured and unlabeled data without human intervention. Deep Learning techniques learn through multiple layers of representation and generate state of the art predictive results. In the past years, Deep Learning techniques have been very successful in performing the sentiment analysis. It provides automatic feature extraction, rich representation capabilities and better performance than traditional feature based techniques. These long-established approaches can yield strong baselines, and their predictive capabilities can be used in conjunction with the arising deep learning methods (Preethi et al., 2017). Two techniques of neural networks are very common - Convolution Neural Networks (CNN) for image processing and Recurrent Neural Networks (RNN) - for natural language processing (NLP) tasks (Goularas & Kamis, 2019). Deep Learning is used to optimize the recommendations depending on the sentiment analysis performed on the different reviews, which are taken from different social networking sites. The Experiments performed indicate that the RNN based Deep-learning Sentiment Analysis (RDSA) improvises the behavior by increasing the accuracy of the sentiment analysis, which in turn yields better recommendations to the user and thus helps to identify a particular position as per the requirement of the user need(Preethi et al., 2017). In this article, we will discuss popular deep learning models which are increasingly applied in the sentiment analysis including CNN, RNN, and various ensemble techniques. This article provides insights on various techniques for sentiment analysis.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VII July 2021- Available at www.ijraset.com

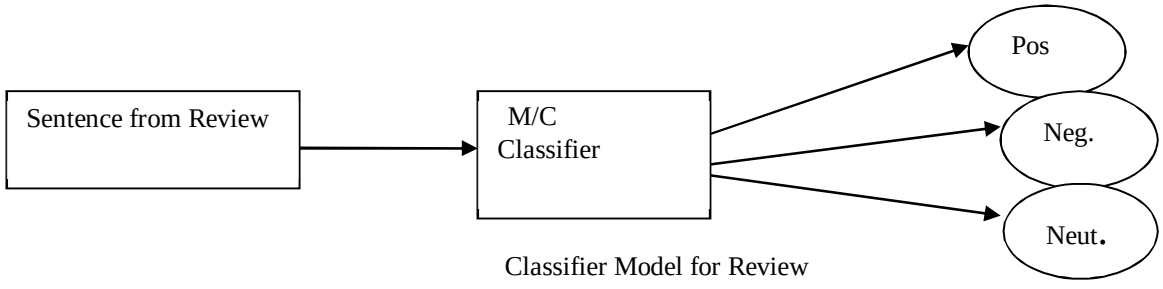
III. MACHINE LEARNING CLASSIFIER FOR SENTIMENTAL ANALYSIS

Machine learning approach is one of the most noticeable methods picking up enthusiasm of scientists because of its flexibility and precision. In assumption examination, generally the managed learning variations of this strategy are utilized. It includes three phases: Data gathering, Pre-handling, Training information, Classification and plotting results. In the preparation information, an accumulation of labeled corpora is given. The Classifier is introduced a progression of highlight vectors from the past information. A model is made dependent on the preparation informational collection which is utilized over the new/concealed content for characterization reason. In machine learning strategy, the way to exactness of a classifier is the determination of suitable highlights. For the most part, unigrams (single word phrases), bi-grams (two continuous expressions), tri-grams (three successive expressions) are chosen as highlight vectors. A work on Chinese OM, which emphasized mining opinions on online reviews, was presented by Zhang et al., [15].This paper was based on machine learning methods. Using a real-world Amazon CN dataset on opinions, it conducted comparative experimental studies and concluded that the approach was effective. Though machine learning based method outperformed its alternatives (especially SVM based method), such methods need large labeled training instances, which are time consuming and labor intensive to get.A classification scheme of pre-release movie popularity using C4.5 and PART classifier algorithm was proposed by Asad et al., [16] which defined relation between post release movies attributes using correlation coefficient. Movie data across the internet makes it a good candidate for machine learning and knowledge discovery. But, most research is directed to bi-polar classification of movie or a movie recommendation system based on reviews by viewers on various internet sites. There is an assortment of proposed includes number of positive words, number of negative words, length of the report, Support Vector Machines (SVM) [24], and Naïve Bayes (NB) calculation [5]. Exactness is accounted for to change from 63% to 80% contingent on the mix of different features chosen. The outcomes have demonstrated that machine learning computations work honorably on weighted unigrams and SVM has come about most prominent precision [12].

Machine learning approach follows the given steps below:

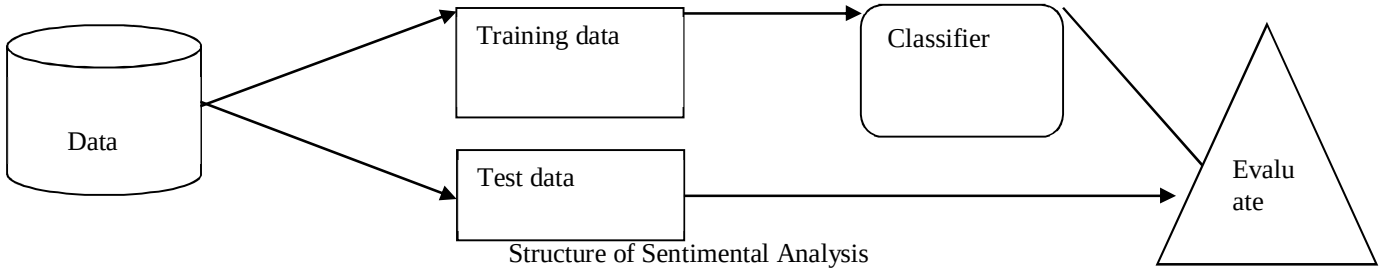
1) *Data Collecting:* In this Initially, experimental data is collected from any ecommerce website, and then it is in csv format
2) *Pre-handling:* In this stage, the gained information is cleaned and prepared for bolstering it into the classifier. Data is preprocessed to remove grammar by the help of data mining preprocessing techniques and remove different grammar from the sentence and like lowercase and uppercase letter.
3) *Training Data:* A hand-labeled accumulation of information is set up by most normally utilized publicly supporting technique. This information is the fuel for the classifier; it will be nourished to the calculation for learning reason. Feature selection is performed to extract relevant features from data set. We only want product name time and product review and in this phase sentiment orientation is determined, Positive and negative tags are append to dataset to each review to conduct supervised learning.
4) *Classification:* This is the core of the entire procedure. Depending upon the prerequisite of the application different Machine Learning classifier like Naïve Bias, SVM, Decision Tree, KNN, Deep forest, are used for the extraction.
5) *Results:* Results are plotted dependent on the sort of portrayal chosen for example diagrams, charts, and so forth. Execution tuning is done preceding the arrival of the calculation. The accuracy is calculated.

1.



Classifier Model for Review

2.

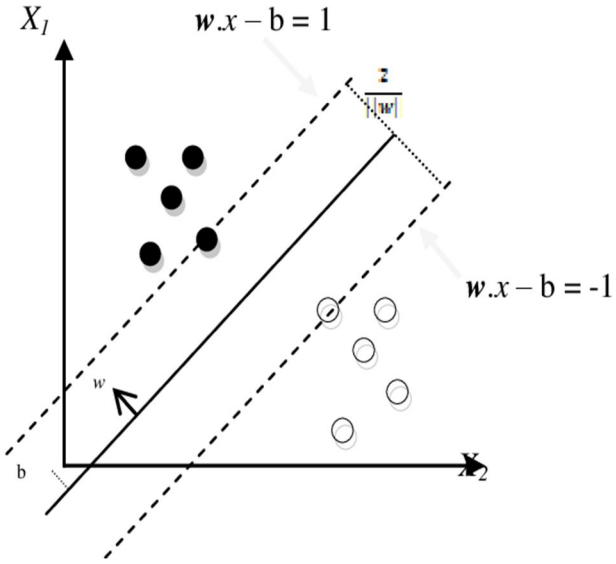


Structure of Sentimental Analysis

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VII July 2021- Available at www.ijraset.com

A. SVM Classifier

SVM is a **supervised (feed-me)** machine learning algorithm that can be used for both classification and regression challenges. Classification is predicting a label/group and Regression is predicting a continuous value. SVM performs classification by finding the hyper-plane that differentiates the classes we plotted in n-dimensional space.



B. Decision Tree Classifier

Decision trees are supervised methods, so they need to be trained on some annotated data. Thus the general idea is the same as for any text classification: given a set of documents (for instance represented as TFIDF vectors) together with their labels, the algorithm will calculate which how much each word correlates with a particular label.For instance it might find that the word "excellent" often appears in documents labeled as positive, whereas the word "terrible" mostly appears in negative documents. By combining all such observations it builds a model able to assign a label to any document.

C. Naive Bias Classifier

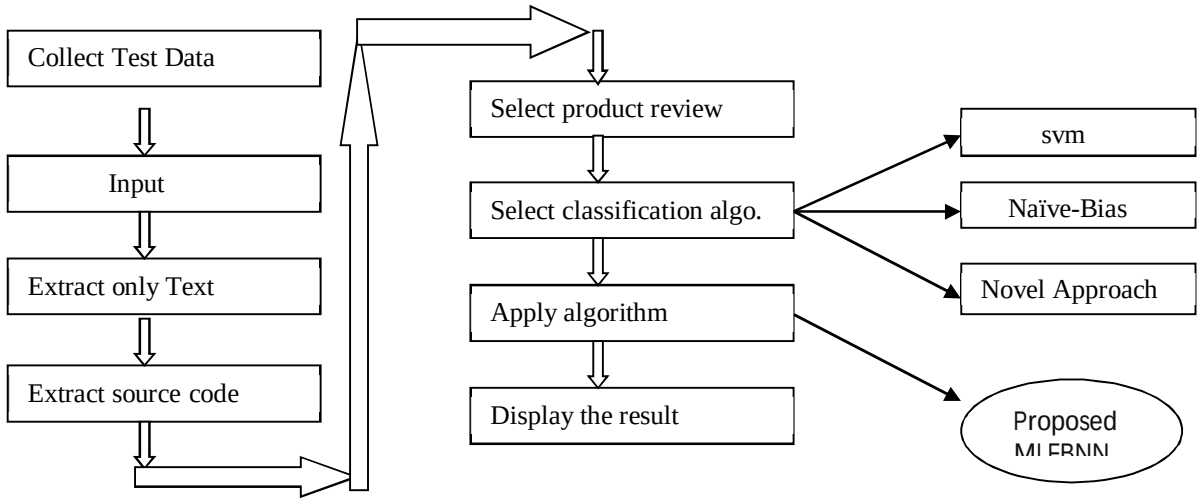
A statistical classifier that maps input feature vectors to output class labels [21]. For a set of training data D, each row is represented by an n-dimensional feature vector, X =x1, x2… xn. There are K classes, K1, K2, ..., Km in the output class label. For every tuple X, the classifier will predict 2 as given by Eq.2 that X belongs to Ki if and only if: P (Ki|X) > P (Kj |X),where i, j = [1,m] and i ! j. P(Ki|X)=n a=1P(xk|Ki)

IV. PROPOSED APPROACH FOR SENTIMENT ANALYSIS

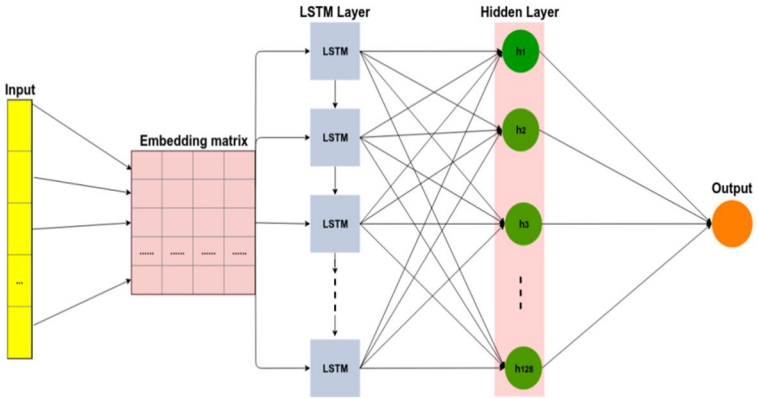
This paper based on mining review from the Websites like amazon.com, which allows user to freely write the view. Sentiment analysis is one of the most rapidly growing research topics in computer science, making it difficult to keep up with all of the activity. Vast number of techniques have grown to analyze the customer reviews. Product reviews plays an important role in sentiment analysis and opinion mining, not only in these process but also in manufacturing field. According to user reviews existing techniques could not accurately detect the product aspect and non-aspect by using the frequent noun method. A novel Algorithm is proposed to detect the product aspect term from review sentence in used with POS tags for sentence segmentation. Term frequency holds the counts of each word and inverse sentence frequency holds counts of words occurs in all sentences and for aspect term we use Feedback Neural Network by this we seprate the word with the sentence.After that we use the polarity method for sentence segmentation like whether it is positive or negative.

Segmented terms are grouped with product aspect in which Feedback Neural Network is used to connect with product aspect feedback loop. Pre-processing of text is also processed in user reviews for removing mistyped words, stop words and company names. Grouped reviews needs to be classified so that opinion of the customer about the product could be identified. Existing techniques uses opinion mining to classify the reviews based on the content but could not provide effective result since decision making problem occurs by classifying it as positive, negative and neutral. The Model is an influential model in language modeling because it doesn't represent the context of fixed-length that contaminate all history words.

The old model uses algorithm such as Naïve Bayes classifier, SVM and Novel SVM algorithm to classify the review as positive and negative review. In recent years, with the further development of natural language research, the traditional feature extraction method based on machine learning has been unable to meet the current demand some scholars tried to use the method of deep learning to solve some problems in Natural Language Processing and achieved good results. For this reason, we also use deep learning model to deal with text sentiment classification problem. In the proposed method following preprocessing task has been completed to classify sentiment analysis from reviews



In this figure, input data is preprocessed to reshape the data for the embedding matrix, next layer is the LSTM and the final layer is fully connected layer for text classification (Dang et al., 2020). The dataset consists of reviews and product information from Amazon was collected. This dataset includes reviews (ratings, text, review) and product metadata (descriptions, category information, price, brand, and image features).



1) Collection of Dataset: The dataset is collected from Amazon and it is in js on format. Each is on file contains number of reviews. Dataset has reviews of Camera, Laptops, Mobile phones, tablets, TVs, video surveillance.
2) Preprocessing: In preprocessing tokenization, stop word removal, stemming, punctuation marks removal, etc., has done. It has converted in bag of words. Preprocessing is important in sentiment analysis and opinion mining.
3) Score Generation In this step, every sentence has analyzed and calculated sentiment score. To calculate sentiment score dataset has compared with opinion lexicons i.e. 2006 positive words and 4783 negative words and calculated sentiment score for every sentence.
4) Sentiment Classification Using score and different features different machine learning algorithms as applied and different accuracy measurements calculated. Proposed method uses the every machine learning algorithm has three components:

A. Representation
B. Evaluation
C. Optimize

1) Algorithm 1: Data Filtration Algorithm
a) Step 1: Importing both positive and negative files and combining them into single file
b) Step 2: Removal of punctuations and numbers from the dataset
c) Step 3: Output (Filtered data)

2) Algorithm 2: Algorithm for Machine Learning Implementation
1) Step 1: Fetching text paragraph from dataset
2) Step 2: Feature Extraction phase: Extracting words corresponds to adjective, adverb and verb.
3) Step 3: All the positive sentences are labeled as "pos" and all the negative ones are labeled as "neg".
4) Step 4: Most frequent feature vector word is set to 5000 words.
5) Step 5: Random shuffling the dataset for training
6) Step 6: Dividing dataset into 70% training and 30% testing dataset
7) Step 7: Training dataset to classification algorithms
8) Step 8: Save the outputs of step 2, and step 7
9) Step 9: Output (Representation of Accuracy of each model)

3) Algorithm 3: Proposed algorithm to perform Sentiment Analysis
1) Step 1: User Input
2) Step 2: Preprocessing:

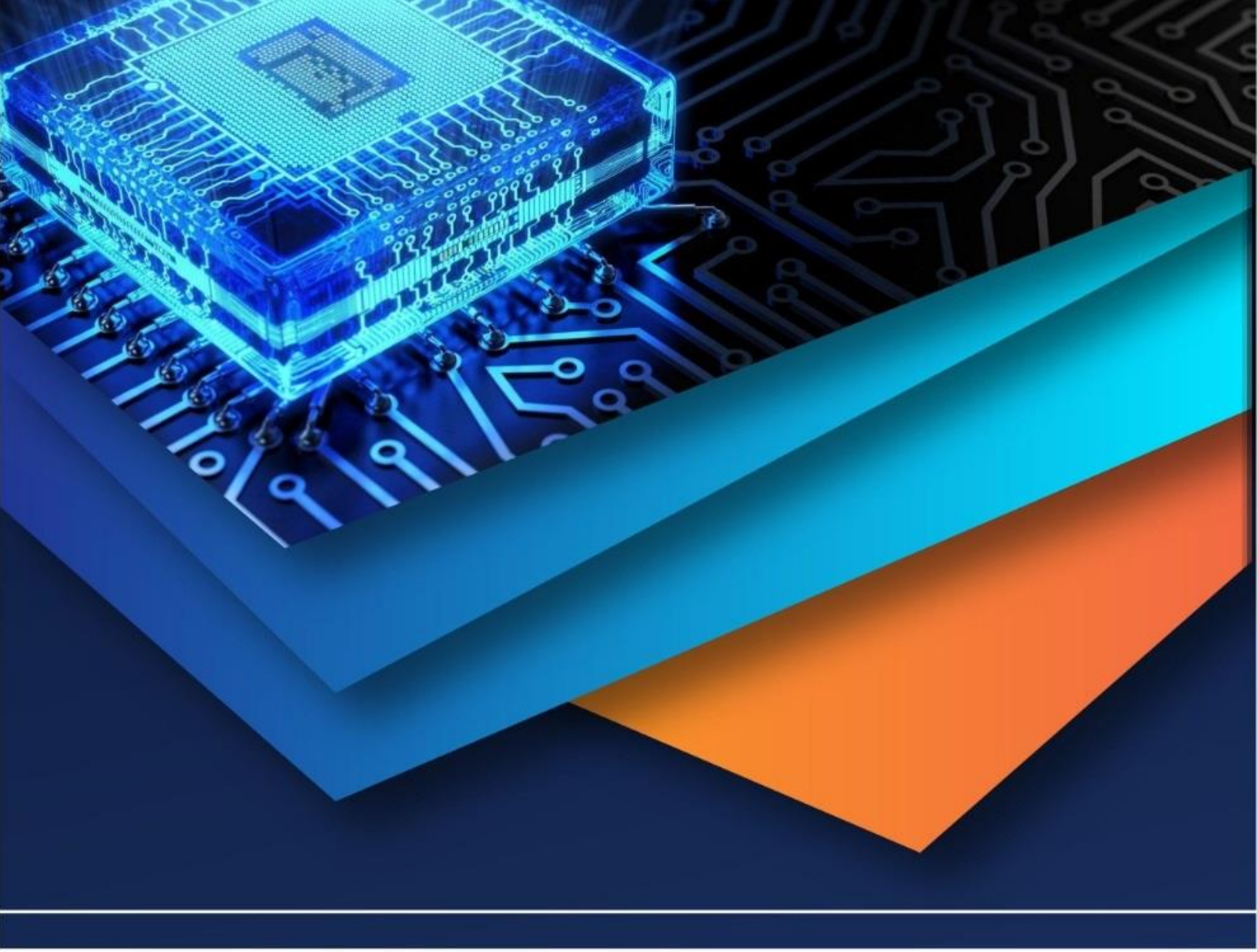
V. CONCLUSION AND FUTURE SCOPE

An evolutionary shift from offline markets to digital markets has increased the dependency of customers on online reviews to a great extent. Online reviews have become a platform for building trust and influencing consumer buying patterns.. Our research is aiming to achieve this by conducting sentiment analysis of product reviews and classifying the reviews into positive and negative sentiment. After balancing the data with almost equal ratio of positive and negative reviews, three classification models have been used to classify Reviews. Out of these classifiers, i.e., Naïve Bayes, SVM and Decision Tree, predictive accuracy of New Proposed RNN is found to be the best. The accuracy results have been increased among the three models. In future, the work can be extended to perform multimodal Sentimental analysis so that the analysis is better when we include Audio Facilities in the review with Text.

REFERENCES

[1] Callen Rain, "Sentiment Analysis in Amazon Reviews Using Probabilistic Machine Learning", Swarthmore College Compter Society, November 2013
[2] P. Russom et al., "Big data analytics," TDWI best practices report, fourth quarter, pp. 1–35, 2011.
[3] S. Erevelles, N. Fukawa, and L. Swayne, "Big data consumer analytics and the transformation of marketing," Journal of Business Research, vol. 69, no. 2, 2016.
[4] V. Hatzivassiloglou and K. R. McKeown, "Predicting the semantic orientation of adjective s," in Proceedings of the eighth conference on European chapter of the Association for Computational Linguistics. Association for Computational Linguistics, 1997, pp. 174–181.
[5] T. Wilson, J. Wiebe, and P. Hoffmann, "Recognizing contextual polarity in phrase-level sentim ent analysis," in Proceedings of the conference on human language technology and empirical methods in natural language processing. Association for Computational Linguistics, 2005, pp. 347–354.
[6] B. Pang and L. Lee, "A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts," in Proceedings of the 42nd annual meeting Association for Computational Linguistics. Association for Computational Linguistics, 2004, p. 271.
[7] A. Pak and P. Paroubek, "Twitter as a corpus for sentiment analysis and opinion mining." in LREc, vol. 10, no. 2010, 2010.
[8] M. WAHYUDI and D. A. KRISTIYANTI, "Sentiment analysis of smartphone product review using support vector machine algorithm-based particle swarm optimization." Journal of Theoretical & Applied Information Technology, vol. 91, no. 1, 2016.
[9] D. N. Devi, C. K. Kumar, and S. Prasad, "A feature based approach for sentiment analysis by using support vector machine," in Advanced Computing (IACC), 2016 IEEE 6th International Conference on. IEEE, 2016, pp. 3–8.
[10] V. Narayanan, I. Arora, and A. Bhatia, "Fast and accurate sentiment classification using an enhanced naive bayes model," in International Conference on Intelligent Data Engineering and Automated Learning. Springer , 2013, pp. 194–201.
[11] S. Tan and J. Zhang, "An empirical study of sentiment analysis for chinese documents," Expert Systems with applications, vol. 34, no. 4, pp. 2622–2629, 2008
[12] Samaneh Moghaddam and Martin Ester, "Opinion Digger: An Unsupervised Opinion Miner from Unstructured Product Reviews", *Proceedings of 19th ACM International Conference on Information and Knowledge Management*, pp. 1825-1828, 2010.

[13] Vidisha M. Pradhan, Jay Vala, Prem Balani, "A Survey on Sentiment Analysis Algorithms for Opinion Mining", International Journal of Computer Applications, Volume 133, No.9, January 2016
[14] G. Sneka, CT. Vidhya, "Algorithms for Opinion Mining and Sentiment Analysis: An Overview", International Journal of Advanced Research in Computer Scinece and Software Engineering, Volume 6, Issue 2, February 2016
[15] Balahur, A., & Montoyo, A. (2008, October). A feature dependent method for opinion mining and classification. In *Natural Language Processing and Knowledge Engineering, 2008. NLP-KE'08. International Conference on* (pp. 1-7). IEEE.
[16] Asad, K. I., Ahmed, T., & Rahman, M. S. (2012, May). Movie popularity classification based on inherent movie attributes using C4. 5, PART and correlation coefficient. In *Informatics, Electronics & Vision (ICIEV), 2012 International Conference on* (pp. 747-752). IEEE.
[17] M. Hagenau, M. Liebmann, and D. Neumann, "Automated news reading: Stock price prediction based on financial news using context-capturing features," Decision Support Systems, vol. 55, no. 3, pp. 685–697, 2013.
[18] T. Xu, Q. Peng, and Y. Cheng, "Identifying the semantic orientation of terms using s-hal for sentiment analysis," Knowledge- Based Systems, vol. 35, pp. 27 9–289, 2012.
[19] I. Maks and P. Vossen, "A lexicon model for deep sentiment analysis and opinion mining applications," Decision Support Systems, vol. 53, no. 4, pp. 68 0–688, 2012.
[20] G. Qiu, X. He, F. Zhang, Y. Shi, J. Bu, and C. Chen, "Dasa:dissatisfaction-oriented advertising based on sentiment analysis," Expert Systems with Applications, vol. 37, no. 9, pp. 6182– 6191, 2010.
[21] T.-K. Fan and C.-H. Chang, "Blogger-centric contextual advertising," Expert systems with applications, vol. 38, no. 3, pp.1777–1788, 2011
[22] Preethi, G., Krishna, P. V, Obaidat, M. S., Saritha, V., & Yenduri, S. (2017). Application of Deep Learning to Sentiment Analysis for recommender system on cloud.
[23] Wang, Z., & Fey, A. M. (2018). Deep learning with convolutional neural network for objective skill evaluation in robot-assisted surgery.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY