iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** IV **Month of publication:** April 2023

DOI: https://doi.org/10.22214/ijraset.2023.50971

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: 7.538
Volume 11 Issue IV Apr 2023- Available at www.ijraset.com

Smart Phone Product Review Using SVM Technique of Sentiment Analysis

Shahida Khan[1], Shivani Arya[2], Riya Upadhyay [3], Kapil Sahu[4]
[1, 2, 4]*Assistant Professor, AITR, Indore*
[3]*Lecturer, Renaissance University*

Abstract: ***Sentiment Analysis (SA) is a task of identifying positive and negative opinions, emotion and evaluation in text available over the social networking sites and the world wide web have been gained quite a popularity in the recent years. The analysis serves as an important feedback for further improvement in the offered services and user experiences. Several techniques have been used recently including machine learning approaches and vocabulary orientated semantic algorithms. This report presents a survey of various techniques and tools have been used in the previous research sentiment analysis process. There is a massive increase in number of people who access various social networking and micro-blogging websites that gives new shapes the impression of today's generation. Many reviews for a specific product, brand, individual, and movies etc. are helpful in directing the perception of people thus the analysts are begun to create algorithms to automate the classification of distinctive reviews on the basis of their polarities in particular : Positive, Negative and Neutral. This machine-driven classification mechanism is referred as Sentiment Analysis. The ultimate aim of this paper is to use support vector machine (SVM) classification technique to classify the feelings of good phone product review that analyses datasets used for classification of sentiments and texts. Also, data sets are used for training as well as testing and implemented through SVM technique for finding the polarity of the ambiguous tweets. The obtain results show to achieve high accuracy as predicted on the basis of reviews of smart phone.***

Keywords: Sentiment Analysis, SVM , Clurstering , Classification, Preprocessing.

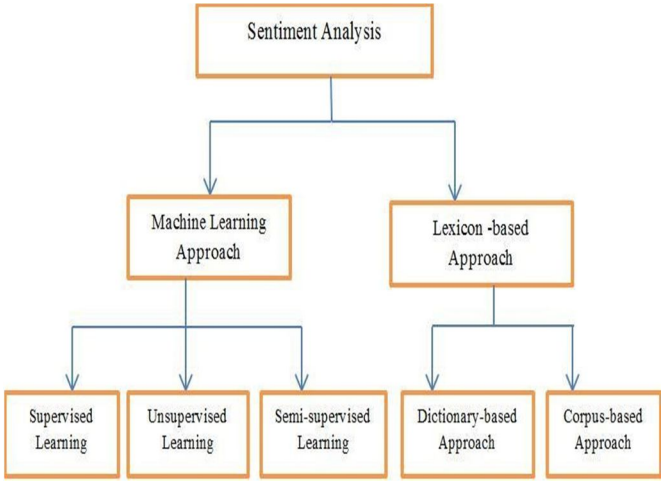
I. INTRODUCTION

The growth of the web and social networking sites such as Facebook, Instagram, Twitter, Blogs, and Forums etc. have been emerged into a huge volume of user reviews and opinions about particular aspects of products or services. People like to share their experiences, thoughts, opinions, feelings, and preferences according to their understanding and observation about the services. Their point of view or impression may be positive, negative or neutral. The opinion is used for identifying trends, user interest, prediction of stock markets, political polls, and market researches, enhancing the user experience by presenting the things of their own interest and to influence them towards a particular direction. For one particular aspect, one may have a positive opinion while some other may have a negative opinion at the same time. Thus, classifying opinion and sentiments of peoples is a difficult task. Furthermore, the shared reviews and feelings are not in specifically structured format, thus identifying its positivity or negativity perspective automatically, is also convenient Therefore, analysis of an unstructured format of text and extract the information for determining the users sentiment requires special machine learning and semantic algorithms for their classification. Sentiment analysis is the process of classifying the opinions conveyed in the documents or statements of the web contents as positive, negative or neutral. The objective of sentiment analysis is mining the opinion behind the users statement and revealing the users interest, preference and thoughts about the particular thing. Various techniques have been presented in the recent years, some rely on the machine learning approaches with supervised, unsupervised or semi-supervised learning and other may use semantic-based approaches. Moreover, few hybrid approaches may also be used from techniques related to different domains. In sentiment analysis, major tasks listed are subjectivity and sentiment classification, sentiment lexicon generation, opinion spam detection and quality of reviews[1].Technically, these machine learning algorithms can be classified into: Supervised, Un-supervised, Semi-Supervised and Reinforcement Learning.

©IJRASET: All Rights are Reserved

If the different input objects are given with a labeled output value (also called supervisory signal) is called supervised machine learning, in contrast to unsupervised learning, where there is no such supervisory signal. between both supervised and unsupervised learning there is semi-supervised learning where there are large amount of input objects and only some of these objects are labeled with output value[2]. However, in Reinforcement Learning there is a learning system to which the training information is provided by the environment based on which it has to discover which action will yield the best reward.Machine learning is a scientific discipline that explores the construction and study of algorithms that can learn from data[12].

Such algorithms operate by building a model supported inputs and victimization that to form predictions or selections, instead of following solely expressly programmed directions Machine learning is closely related to computational statistics. Sentiment analysis techniques are classified into the following categories as shown in fig.1.1



II. LITERATURE REVIEW

In 2016 Fang Luo et al.[4] projected a technique, that would perform at the sentence level and document level whereas it unsuccessful at word level sentiment analysis. In 2016, Ebru Aydogan et al. [1] dole out a comprehensive survey on Sentiment Analysis victimization differing types of Machine learning algorithms and located that SVM and NB were most ordinarily used thanks to their higher estimation capability. prophet Qasem et al. [5] used 2 weighting schemes specifically Unigram term frequency (TF) and written word term frequency- inverse term frequency (TF-IDF) to classify completely different tweets into positive, negative and neutral categories. Positive tweets were determined by positive emoticons and negative tweets by negative emoticons whereas the neutral tweets were outlined as those with no emoticons or keywords that indicate polarity, like happy, sad, good, bad, etc. The liability was thanks to the automated annotation of neutral category. M.Trupthi et al. [6]explored machine learning approaches with completely different feature choice schemes, to spot the simplest attainable approach and located that the classification victimization High data options, resultedin more accuracy than written word Collocation. They additionally projected that there was a scope for improvement victimization hybrid techniques with numerous classification algorithms. Orestes Apple et al. [7] projected a hybrid system victimization Naive mathematician (NB) and most Entropy (ME) to identical dataset thatworkedo.k.withthehighlevelofaccuracyandexactness.S.Brindhaetal.[3] bestowed the survey result on totally different classification techniques (NB, KNN, SVM, DT, and Regression) therefore on notice the classification accuracy for various datasets. They found that just about all classification techniques were suited to the characteristics of text knowledge. They ended that more study on classification development will get the improved quality of text results and correct knowledge at the side of reduced accessing time.In 2015 Huang Zou et al. [8] introduced a grammar feature in pre-existing words- bag technique that unconcealed a lot of on pos tags. They applied SVM and NAVIE mathematician and discovered that word dependency and pos tags did improve the accuracy whereas shut feature did see abundant optimization. In 2014P.Kalaivani and K.L.Shunmuganathan .[9] projected AN improved KNN formula by incorporating data gain for feature choice to enhance and show that this approach outperformed NAVIE mathematician and KNN. They additionally mentioned that they'd be evaluating the model for cross-domain sentiment analysis.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

Mustafa Karamibekr and Ali A. Ghorbani .[10] targeted their work in the main on topic oriented opinion mining, wherever solely opinions concerning specific topic or issue would be thought-about. A text might not essentially contain opinion a couple of targeted topic. As per their experiments the exactness or Recall technique for sound judgment of classification at the sentence level was significantly less than those of previous works. However, the F-measure was higher that indicated AN improved overall balance between exactness and Recall. In 2012Saif et al.[11] worked on linguistics sentiment analysis of twitter and that was projected their results show that the linguistics feature model outperforms the Unigram and POS baseline for characteristics each negative and positivesentiment.Alexander Pak and Saint Patrick Paroubek et al. [12]they worked on small blogging, their analysis paper entitled "Twitter as a Corpus for Sentiment Analysis and Opinion Mining" at their time Twitter was the foremost standard small blogging platform, in order that they targeted thereon solely. In their analysis, they bestowed a technique for AN automatic assortment of a corpus that would be wont to train a sentiment classifier. In 2010 Khin Phyu PhyuShein et al.[13] projected that the mixture of victimization linguistic communication process techniques (NLP), metaphysics supported Formal thought Analysis (FCA) style, and Support Vector Machine (SVM) might used for classifying the package reviews were positive, negative or neutral. Sun Yueheng et al.[14] projected a technique, through that they may decide the mood of userreview. And {to do to try to to|to try And do} this they took And approach for automatic sentiment analysis by: increasing the initial sentiment words by an unvarying method however through this, they may solely bring home the bacon a mean exactness of eightythree.52%.In 2009 Cheng Mingzhi at el.[15] projected a technique within which a word association graph was first of all made from a text corpus, i.e. product reviews, within which every node was a word and if there's a foothold between 2 words, it means that the 2 words co-occur within the same sentence. And then, with a stochastic process formula, the sentiment score was calculated for all the words within the graph at just once. victimization SVM techniques within the real application has some disadvantages, that its terribly troublesome select to settle on to decide on} applicable parameters for the kernel operate and choose the correct feature to construct vectors. They projected a technique within which a word association graph was first of all made from text corpus, i.e. product reviews. within which every node was a word and if there was a foothold between 2 words, it means that the 2 words co-occur within the same sentence. And then, with a stochastic process formula, the sentiment score is calculated for all the words within the graph at just once. victimization SVM in real application has some disadvantages, that its terribly troublesome select to settle onto decide on} applicable parameters for the kernel operate and choose the correct feature to construct vectors. In 2007 Most afa Al Masum Shaikh et al.[16] projected approach to sense sentiments contained during a sentence, by applying a numerical-valence primarily based analysis. For the longer term work, they additionally need to classify texts supported many feeling-types following the OCC emotion model and perform evaluations victimization on-line resources (e.g. blogs, news etc.)

In 2017 Maria et al.[17] projected the utilization of a hybrid approach for the prediction of sentiment, within which Context-sensitive cryptography offered by Word2Vec and sentiment/emotion data offered by a lexicon were combined and this was done to urge the simplest leads to terms of potency, accuracy and also the method time. In 2016 Orestes Apple et al.[18] projected a hybrid system approach to the Sentiment Analysis drawback at the sentence level and provides a high level of accuracy of eighty eight.02% and exactness of eighty four.24% share . Abinash Tripathy et al. [2] used totally different supervised machine learning formula and by applying them he tried to classify flickreviews. IoannisKorkontzelos et al. [19] the y did a research on the impact of sentiment analysis, on extracting adverse drug reactions from tweets and forumposts.Aliaksei Severyn and Alessandro Moschitti [20] they foreseen polarities at each message and phrase level by deep learning approach to sentiment analysis of tweets, and for this, they used an unsupervised neural language model which trained initial word embedding, and further was tuned to seek out the polarities. PreslavNakovand TorstenZesch [21] proposed their paper which presented a task on the evaluation of Compositional Distributional Semantics Models on full sentences organized for the primary time within SemEval-2014. Gokulakrishnan et al. [22] they introduced a paper, during which they analyzed tweets from micro blogging twitter site, and classified them as positive, negative and irrelevant. And further, they studied the performance of varied classifyingalgorithms. Bernard J. Jansen et al. [23] through In their research, they examined eWOM branding. They further analyzed branding comments, sentiments, and opinions in additional than 150,000 microblogpostings.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue IV Apr 2023- Available at www.ijraset.com

III. EXISTING SYSTEM

Nowadays, due to the rapidly growing of e-commerce, more online reviews for products and services are created. Text Mining is used to extract valuable information from large amount of data. A key component is utilized to interface together the extricated data to frame new realities or new theories to be investigated further by more customary method for experimentation.

There are several challenges in Sentiment analysis.

The first is that associate opinion word that's thought-about to be positive in one scenario is also thought-about negative in another scenario. The second challenge is that people don't always express opinions in the same way. The usual text - process depends on the actual fact that tiny variations between 2 items of text don't modification the that means noticeably. Sentiment analysis helps to find words that indicate sentiment and helps to understand the relationship between textual reviews and the consequences of those reviews. The proposed system shows that a simplified version of the sentiment-aware autoregressive model can produce very good accuracy for predicting the box once sale using online review data. It uses document level sentiment analysis that consists of Term frequency and Inverse Document frequency. In this process, mining techniques are applied on online product reviews and predict the collection of the cell phone based on the reviews and analyses how much effect the reviews have on company collection. company collection for the next day is predicted based on online reviews of the present day. A prediction of high or low assortment is additionally foretold.

IV. PROBLME STATEMENT

The existing system uses Navy Bayes Algorithm with parts of discourse labels. Estimation examination is the procedure used to decide the state of mind/supposition/feeling communicated by a man around a specific theme. Notion examination or assessment mining uses normal dialect preparing and message investigation to distinguish and separate subjective data in source materials. The ascent of online networking, for example, websites and informal communities has fuelled enthusiasm for supposition examination. With a specific end goal to recognize the new open doors and to deal with the notorieties, agents typically see the audits/appraisals/proposals and different types of online supposition. Record level fee ling examination means ordering the general estimations communicated by the creator in the entire archive content in positive, negative or impartial classes. A perspective based sentiment surveying framework takes as information an arrangement of literary audits and some predefined viewpoints, and distinguishes the extremity of every angle from every survey to deliver an assessment survey. The notion examination is regularly performed on one single level, for example, substance level, sentence - level, and record level. In substance level vocabulary is fabricate and after that by recognizing earlier and relevant extremity critical components are separated in view of that elements assumption examination is performed. At sentence-level and record level, reports are ordered by general notions, yet not by subject. Classification accuracy isless.

V. SENTIMENT ANALYSIS OFSMART

A. Phone Product Re-View Using SVM Classification Technique

In this section, the proposed methodology includes the subsequent key points: The proposed methodology discusses about applying sentiment analysis and machine learning techniques to review the relationships among the web reviews for smart phone products and therefore the revenue of performance. The techniques are applied on online product reviews and predict collection of the merchandise supported the reviews and analyses what proportion effect the reviews wear the gathering. the merchandise collection for subsequent day is predicted supported online reviews of this day. A prediction of high or low collection is additionally predicted. From the web site, the detailed information containing the values for the following: brands, product date, rank of sale, users reviews, etc. of a sensible phone were obtained. a part of Speech (POS) model during which a sentiment or textual review is represented as a vector, whose entries correspond to individual terms of a vocabulary. Part-of-speech information is meant to be a big indicator of sentiment expression. The score of every sentence within the dataset is calculated by sum of weight of every term within the corresponding sentences. Clustering of the review data supported the TF-IDF measure has been performed. Finally the proposed work shows that accuracy is very achieved, the reviews are taken as appropriate and therefore the success or failure of the smart phone product is predicted supported the reviews. the entire work is implemented in java by using following modules preparing as shown in Fig. 3.1 the entiremethod-

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue IV Apr 2023- Available at www.ijraset.com

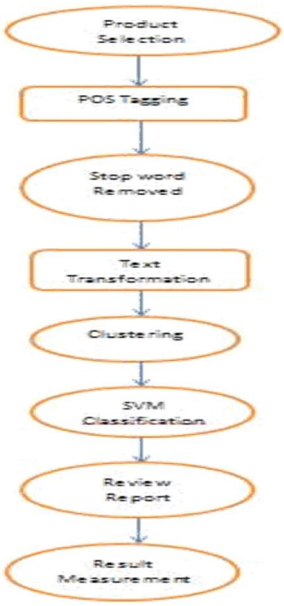


Figure 3.1: Flow of proposed methodology has been consisting variousphases.

B. Text Pre-Processing

Text pre-processing techniques are divided into two subcategories which are POS tagging and stop words removal. In POS, textual data comprises block of characters called tokens. The input reviews are separated as tokens and begin the pre-processing. A stop-list is that the name commonly given to a group or list of stop words. a number of the more frequently used stop words for English include a, of, the, I, it, you, and these are generally considered functional words which don't carry meaning. Hence remove those words that support no information for thetask.

C. Transformation

In the transformation process, the score for every sentence is calculated within the document. For that, first the load of every term is calculated by the merchandise of term frequency and inverse documentfrequency.

D. Clustering

Clustering of the document review is predicated on the TF-IDF measurement. Thus, points on the sting of a cluster, maybe within the cluster to a lesser degree than points within the center of cluster. It chooses the amount of clusters and it findscentroid.

E. SVM Classification

After the removal of the outliers supported the clustering, the improved feature sets were used for sentiment classification. SVM is especially used for the sentiment classification. It classifies the positive and negative reviews.

VI. PERFORMANCE EVALUATION

The performance of algorithms depends on various parameters like Precision, Recall, F-Measure and Accuracy. this will be understood using True Positives, False Positives, True Negative and False Negative, where, True Positive (TP) is items correctly labeled as belonging to the positive class False Positives (FP) are items incorrectly labeled as belonging to the positive class True Negative (TN) are items correctly labeled as belonging to the negative class False Negative (FN) are items incorrectly labeled as belonging to the negativeclass.

In information retrieval, precision may be a measure of relevant instances among the retrieved instances; it's the ratio of number of elements correctly labeled as positive to total number of positively classified elements3.2.

$$Precision = T \quad P=(T \quad P + F \quad P \quad) \tag{3.1}$$

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

Whereas, Recall may be a measure of what percentage truly relevant results are returned; it's the ratio of total number of positively labeled elements to total elements which are truly positive3.2.

$$Recall = T P=(T P + F N) \quad (3.2)$$

F-Measure may be a measure that mixes precision andrecall andis that the mean of precision and recall3.3.

$$F Measure = 2(Precion Recall)=(Precision + Recall) \quad (3.3)$$

Accuracy is additionally used as a statistical measure of how well a classification test correctly identifies or excludes a condition. That is, the accuracy is that the proportion of true results (both true positives and true negatives) among the entire number of cases examined3.4.

$$Accuracy = (T P + T N) = (T P + T N + F P + F N) \quad (3.4)$$

Accuracy may be a common measure for classification performance. it's the proportion of correctly classified examples to the entire number of examples, while error rate uses incorrectly classified rather than correctly[1].

VII. CONCLUSION

This chapter is that the concluding a part of the dissertation work and also proposes some suggestions towards which this work are often further extended. Section 4.1 brings out the general conclusions of the research work administered during this thesis. Section 4.2 gives the longer term research directions and possible extensions of the work presented with thesis Sentiment analysis is one among the recent research area now a days. the knowledge gathered from the info sources like blogs, forums, review sites etc. has been playing a crucial role in expressing people's feelings, thoughts, emotions, and opinions for the actual issue or topic. The proposed show works on gathering of tweets identified with smart phone reviews. The exactness has enhanced in differentiation to the various mixes of models utilized by scientists already. The outcomes produce 90.99 % accuracy. Therefore, it are often derived that sentiment analysis is improved by using Support Vector Machine (svm). It work well with predefined quite sentence which we've indicated.

REFERENCES

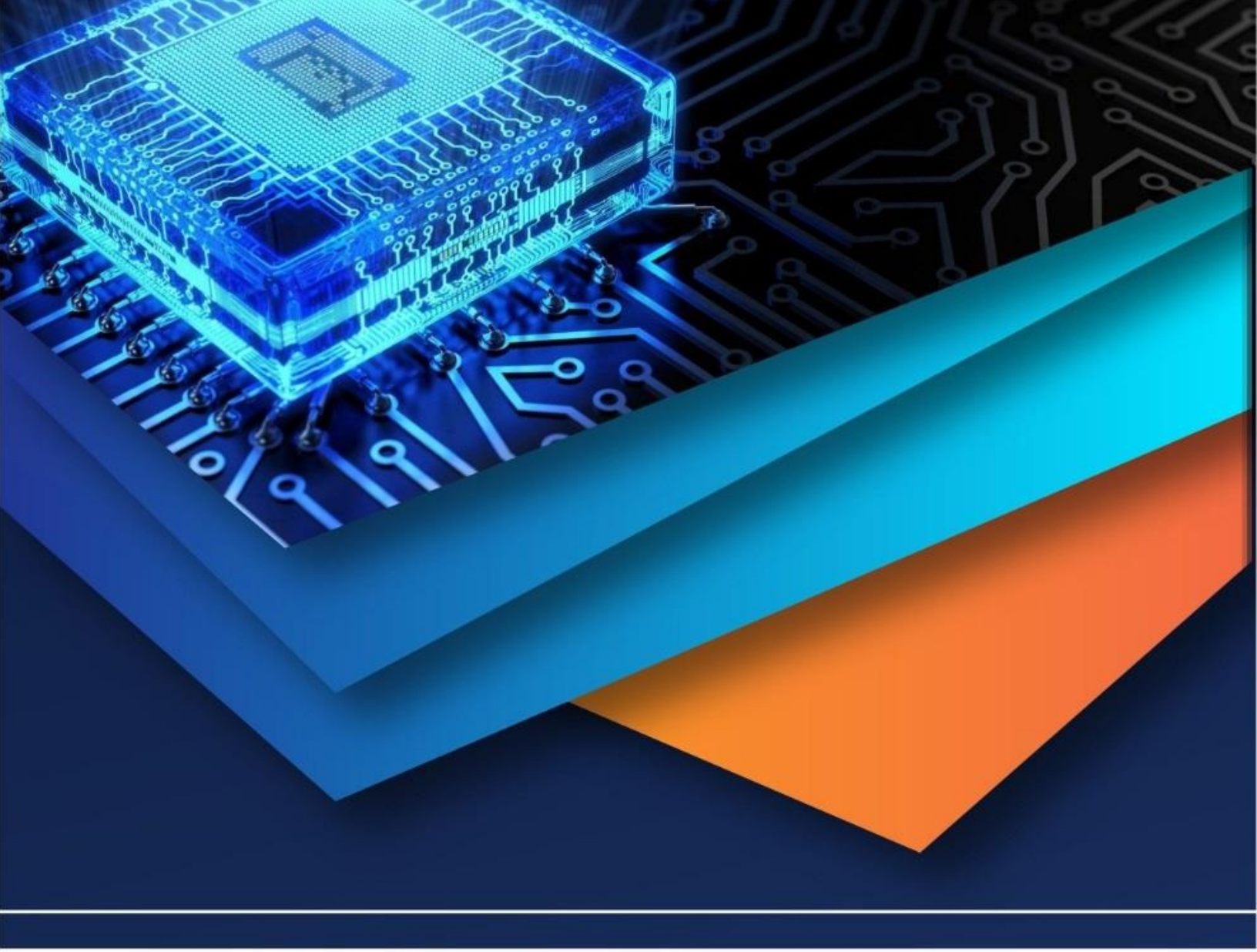
[1] Ebru Aydo˘gan and M Ali Akcayol. A comprehensive survey for sentiment analysis tasks using machine learning techniques. In INnovations in Intelligent SysTems and Applications (INISTA), 2016 International Symposium on, pages 1–7. IEEE,2016.
[2] Abinash Tripathy, Ankit Agrawal, and Santanu Kumar Rath. Classification of sen-timent reviews using n-gram machine learning approach. Expert Systems with Ap-plications, 57:117–126,2016.
[3] S Brindha, K Prabha, and S Sukumaran. A survey on classification techniques for text mining. In Advanced Computing and Communication Systems (ICACCS), 2016 3rd International Conference on, volume 1, pages 1–5. IEEE, 2016.
[4] Fang Luo, Cheng Li, and Zehui Cao. Affective-feature-based sentiment analysis usingsvm classifier. In Computer Supported Cooperative add Design (CSCWD), 2016 IEEE 20th International Conference on, pages 276–281. IEEE,2016.
[5] Mohammed Qasem, RuppaThulasiram, and ParimalaThulasiram. Twitter sentiment classification using machine learning techniques for stock markets. In Advances in Computing, Communications and Informatics (ICACCI), 2015 International Conference on, pages 834–840. IEEE,2015.
[6] M Trupthi, Suresh Pabboju, and G Narasimha. Improved feature extraction and classification sentiment analysis. In Advances in Human Machine Interaction (HMI), 2016 International Conference on, pages 1–6. IEEE, 2016.
[7] Orestes Appel, Francisco Chiclana, Jenny Carter, and Hamido Fujita. A hybrid approach to sentiment analysis.2016.
[8] Huang Zou, Xinhua Tang, Bin Xie, and Bing Liu. Sentiment classification using machine learning techniques with syntax features. In Computational Science and Computational Intelligence (CSCI), 2015 International Conference on, pages 175– 179. IEEE,2015.
[9] P Kalaivani and KL Shunmuganathan. An improved k-nearest-neighbor algorithm using genetic algorithm for sentiment classification. In Circuit, Power and Computing Technologies (ICCPCT), 2014 International Conference on, pages 1647–1651. IEEE, 2014.
[10] Mostafa Karamibekr and Ali AGhorbani. A structure for opinion in social domains. In Social Computing (SocialCom), 2013 International Conference on, pages 264–271. IEEE,2013.
[11] Hassan Saif, Yulan He, and Harith Alani. Semantic sentiment analysis of twitter. In International Semantic Web Conference, pages 508–524. Springer,2012.
[12] Alexander Pak and Patrick Paroubek. Twitter as a corpus for sentiment analysis and opinion mining. In LREc, volume 10, pages 1320–1326, 2010.
[13] Khin Phyu PhyuShein andThiThi Soe Nyunt. Sentiment classification supported ontology and svm classifier. In Communication Software and Networks, 2010. ICCSN'10. Second International Conference on, pages 169–172. IEEE,2010.
[14] Yueheng Sun, Linmei Wang, and Zheng Deng. Automatic sentiment analysis for web user reviews. In informatics and Engineering (ICISE),

2009 1st Inter-national Conference on, pages 806–809. IEEE,2009.

[15] Cheng Mingzhi, Xin Yang, Bao Jingbing, Wang Cong, and Yang Yixian. A stochastic process method for sentiment classification. In Future Information Technology and Management Engineering, 2009. FITME'09. Second International Conference on, pages 327–330. IEEE,2009.

[16] Mostafa Al Masum Shaikh, Helmut Prendinger, and Mitsuru Ishizuka. An analytical approach to assess sentiment of text. In Computer and knowledge technology, 2007. iccit 2007. 10th international conference on, pages 1–6. IEEE, 2007.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



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)