iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** IV **Month of publication:** April 2022

DOI: https://doi.org/10.22214/ijraset.2022.42005

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 10 Issue IV Apr 2022- Available at www.ijraset.com

A Survey on Sentimental Analysis Approaches using Machine Learning Algorithms

Anith Ashok[1], Dr. Sandeep Monga[2]
[1, 2]*Oriental Institute of Science and Technology*

Abstract: ***Sentiment analysis is the measurement of positive and negative language. It is a way to evaluate written or spoken language to determine if the expression is favourable unfavourable, or neutral, and to what degree. It is one of the prominent fields of data mining that deals with the identification and analysis of sentimental contents generally available at social media. In this work, a survey has been conducted on various work done in the past on sentiment analysis which includes opinion mining methods, machine learning based approaches and hybrid approaches which combines the both. Some approaches which uses special statistical and machine learning models are included in separate section. It is concluded from the work that sentiment analysis on text can be improved if combined with other features like emojis. It can further be improved by using hybrid models.***

Keywords: Time series forecasting, Sentiment Analysis, Hybrid Models

I. INTRODUCTION

The unexampled increase in the acceptance as well as penetration of social media platforms, similar as Facebook, Twitter, Google plus,etc., in a day to day life, have changed the pattern of online communication of people. Formally, user's online access was largely confined to professional contents similar as news agencies or corporations. Still, these days they can seamlessly interact with each other in a further concurrent way by creating their own content within a network of peers. According to Howard [1], " We use Facebook to record the challenge, Twitter to coordinate, and YouTube to tell the word".

Social media has surfaced as a vital platform of representing people's sentiment, boosting the conditions of data mining in the field of the sentiment analysis. In the sentiment analysis, the raw data is the online text that's changed by users through social media [2]. Twitter, which is one of similar social media, has come the prominent source to change the online text, furnishing a vast platform of sentiment analysis. Twitter is a veritably popular social networking website that allows registered users to post short dispatches, also called tweets, up to 140 characters.

Twitter database is one of the largest database having 200 million users who post 400 million communications/tweets in a day [3]. At Twitter, users frequently partake their particular opinion on different subjects similar as acceptance or rejection of politicians and standpoint about products, talk about current issues and partake their particular life events. Still, users post their tweets with smaller characters by using a short form of words and symbols similar as emoji. Thus, analysis of these tweets can be used to find strong shoes and sentiments for any content.

Twitter data has formerly been used by different people to prognosticate stock request vaticination [4], box office earnings for pictures [5], identify the guests with negative sentiments [6]. Sentiment analysis (SA) Sentiment analysis is the dimension of positive and negative language. It's a way to estimate written or spoken language to determine if the expression is favourable, unfavourable, or neutral, and to what degree. It's one of the prominent fields of data mining that deals with the identification and analysis of novelettish contents generally available at social media. Moment's algorithm- grounded sentiment analysis tools can handle huge volumes of client feedback constantly and directly. Paired with textbook analytics, sentiment analysis reveals the client's opinion about motifs ranging from your products and services to your position, your announcements, or indeed your challengers. Why is sentiment analysis important? Sentiment analysis is critical because helps you see what guests like and dislike about you and your brand. Client feedback from social media, your website, your call centre agents, or any other source contains a treasure trove of useful business information. But, it is n't enough to know what guests are talking about. You must also know how they feel. Sentiment analysis is one way to uncover those passions. The main end of sentiment analysis is to determine the station of users on a particular content. Sentiment analysis styles can be colossally distributed into lexicon- grounded styles, machine literacy-grounded styles, and mongrel styles. In this work, a sentiment analysis approach for twitter data grounded on ensemble literacy styles is proposed.

©IJRASET: All Rights are Reserved

II. LITERATURE REVIEW

Sentiment analysis identifies the opinion or polarity of reviews. For example, it may be the writer's attitude on movie, book, recipe, product, etc., which may be positive, neutral or negative. In short, the sentiment analysis is based on people's opinions and attitudes. The accuracy of automatic sentiment analysis should agree with the judgments of humans. Various machine learning algorithms used in the sentiment analysis for classification of sentiments are discussed in this section. In this modern world, the internet is used commercially in all businesses by means of websites, online portals, reviews, feedback, recommendations, and blogs. Many viewers write their opinions on different products they use in real time through social media. The reviewers provide their opinion on different aspects such as book, hotels, products, movies, research, events, etc. There are many challenges in the analysis of sentiment and these challenges need to be addressed, while analysing the performance of sentiment analysis and detecting the sentiment polarity on any product.

A. Opinion Mining Methods

Sentiment mining is also referred to as opinion mining. Bo Pang et al. (2002) [7] considered the problem of classifying documents not by topic, but by overall sentiment, e.g., determining whether a review is positive or negative. Using movie reviews as data, standard machine learning techniques definitively outperform human-produced baselines. However, the three machine learning methods employed (Naive Bayes, maximum entropy classification, and support vector machines) do not perform well on sentiment classification with traditional topic-based categorization. This paper is concluded by examining factors that make the sentiment classification problem more challenging. David M. Beli et al. (2003) [8] proposed a generative model for text and other collection of discrete data that aims to improves on several previous probabilistic latent semantic indexing models. In the context of text modeling, the model proposed posits that each document is generated as a mixture of topics, where the continuous-valued mixture proportions are distributed as a latent Dirichlet random variable. It shows good results on applications of this model to problems in text modeling and classification.

Krishnakumari et al. (2007) [9] proposed sentiment classifiers with domain adaptation, focusing on online reviews for different types of products. First it works to reduce the relative errors due to adaptation between different domains over the actual Structural Correspondence Learning (SCL) algorithm. Next it identifies a measure of domain similarity that relates to the adaptation of a classifier from one domain to another. This could be used as an instance to select a subset of domains to train classifiers to form one domain to another. Daniel Ramage et al. (2009) [8] introduced Labeled LDA, a topic model that controls Latent Dirichlet Allocation with a one–to–one correspondence between latent topics of LDA. This makes Labeled LDA learn directly from a tag of words. The results show improved performance over traditional LDA methods. Since it is a multilabel text classifier, it is comparatively better than discriminative baseline models on a variety of datasets. Taboada et al. (2011) [10] have proposed the machine learning methods to Support Vector Machine (SVM) classifiers. It used to calculate the high polarity classification value, which is used for training specific datasets. It is a supervised method that aims to discover patterns with vector representations of texts in natural language. The collected texts are optimized, but the model requires more training data and training time to reach this performance level. Liu and Chen (2015) [11] discovered different multi-label classifiers for sentiment classification. It was proposed with eleven multilevel classification methods, examined on two micro-blog datasets and considered eight different evaluation metrics for analysis. The method also used three different sentiment dictionaries for multi-level classification. With the view of authors, the multi- label classification process is done in two stages: problem transformation and algorithm adaptation. In the problem transformation phase, the problem is transformed into multiple single-label problems. In the training phase, the system learns from these transformed single label data and in the testing phase it makes predicts a single label and translates it into multiple labels. The data is transformed as per the requirement of the algorithm in the algorithm adaption process. Norambuena et al. (2019) [12] has implemented the sentiment analysis with the opinion methods that can perform on several levels as document and sentence levels. The document level is analysed whether the document expresses positive or negative sentiments. The sentence level analyses the sentence by word sentiment and it determines the statement fact. The adjacency provides a way to identify the features of the product used in the first attempt because it is essential in the frequent pattern mining algorithms.

Rajiv Bajpai (2019) [13] has introduced an Aspect-Sentiment embedding using the opinion mining methods, which helped to overcome the drawbacks of a company product that has a negative review given by the customer of the company. In this method, there are two types of aspects referred to as implicit and explicit. Explicit are the observable behaviours, rituals, symbols and heroes of culture. The implicit methods have underlying values guided by people regarding any aspect in the real world and provide suggestions as appropriate or inappropriate. The method works to assign polarity to the aspects and calculate the score of an aspect belonging to the reviews of the company.

The dataset used in this method is Glassdoor source with 40k reviews. Fabio Garza (2018) illustrated a methodology to reach the objective with reviews in the field of public security. The emotion is analysed for indicating the most meaningful words in the text of primary emotions. The perception of risk is considered and the keywords were chosen from the mental viewpoints, which gave the impression of the positive and negative feelings on a context. Catchphrases analysis is made utilizing appropriate parameters to quantify how significant the word or word pair is informative in the content of the document. It is realized through this work that vocabularies for the Italianlanguage is helpful for this sort of investigation. It is now accessible, yet a few words are explicit for the unique circumstance and it needs to creation of another dictionary. Omkar Sunil (2018) [14] introduced an open source tool for sentiment analysis. It does not concentrate on the basic opinion method to get into the profound dimensions of grouping that uncover concealed notions behind the information posted by the customers. This additionally aims to conquer the majority of the difficulties referenced in the existing ones by mapping different Application Programming Interfaces (APIs). It can be used to perceive the market estimation of the business brand and grasp the general attractive quality of the item according to comments posted by users. It gathers the data about the trends, emotions, dislikes, attitudes, likes. The information from the geographic locations makes a trend in the brand of the product. The sentiment analysis has more benefits in this method of cloud and software.

B. Machine Learning Based Approaches

Manual analysis of discovering sentiments and opinions in a large volume of textual data is extremely difficult. So, in recent years, there have been interests in the natural language processing community to develop novel text mining techniques with the capability of accurately extracting customers' opinions from large volumes of unstructured text data (Lin and He 2009) [15]. An approach to classify texts as positive or negative using Support Vector Machines (SVMs), a well-known and powerful tool for classification of vectors of real-valued features was discussed by Zhang (2012) [11] with its application. Sentiment Analysis assignment can be explained as an arrangement task where every classification speaks to an assumption. This gives organizations, a way to assess the degree of item acknowledgment and to decide methodologies to improve item quality. It also encourages producers or government officials to investigate open suppositions concerning strategies, open administrations or political issues. Nong Ye et al. (2002) [16] focused on incremental learning algorithms. This paper presents a new data mining algorithm, Clustering and Classification Algorithm (CCA) for clustering and classification. It is based on supervised clustering and instance-based classification. It is used for many classification problems, like statistical quality control to detect inconsistent patterns in the manufacturing process, group technology, shop floor control, and so on. These are places where computers are used mainly to collect large amounts of process data. Riverola et al. (2007) [17] introduced two complementary techniques which could select both terms and e-mails representatives of the current situation. This proposed system is evaluated against other well-known successful lazy learning approaches within a cost-sensitive framework. As the first improvement it showed that it can handle the concept drift inherent in e-mail spam data, which allows easy updating of new types of spam as they arrive. Next the instance-based approach for filtering the spams allows the sharing of instances with the effort of labeling e-mail as spam. Ding et al. (2008) [18] focused on customer reviews of products. In particular, the author reviewed the problem of determining the semantic of opinions that are expressed on the features of products in reviews. The authors proposed a universal approach that can accurately detect the semantic orientation of the sentiment word based on the context of the review. It proposes a new function which is used to combine multiple attitude words in the same sentence. Liang et al. (2009) [19] decide to focus on the online learning problems. The work proposed an incremental learning method for a nonlinear PSVM (Proximal-SVM) classifier to enable online learning in the PSVM classifier and make it efficient for classification. Mathematical analysis and the obtained results gave an indication indicating that these methods can reduce computation time without degradation in accuracy. Hamed Malek et al. (2011) [20] developed three new learning algorithms for fuzzy systems based on training errorand genetic algorithm. The first two algorithms have two phases. In the first phase, the optimum points of training data in input-output space are used to create the initial structure of the neuro fuzzy network with k–Nearest Neighbour and Mean-Shift methods. It iteratively adds new neurons based on an error-based algorithm. In the second phase genetic algorithm is used to remove the repeated neurons. The third algorithm constructs One R with a modified version of the error algorithm that is used in the first two methods. The algorithms are simple and have low computational costs. They are effective in the approximation of nonlinear functions with good accuracy and minimal rules. Sotiris Kotsiantis (2013) [21] focuses on Incremental Learning. This paper proposed an incremental learning method using the NB Classifier and k-NN Classifier. To increase the prediction accuracy, the NB Classifier is integrated with the k-NN Classifier. The proposed method is compared with other algorithms on several datasets which produced better accuracy in most of the cases.

Yukun Ma, (2018) [22] has implemented The Aspect Based Sentiment Analysis (ABSA) by using the Long Short-Term Memory (LSTM) network. It is proposed to show hierarchical attention to explicit the first target on the entire sentences and to extend the classic LSTM components for integrating the external knowledge. ABSA is a task that classifies the sentiment polarity concerning the aspects. It can represent the information in the sentence in the analysis, the benefits of the analysis are to overcomes the shortness of recurrent neural network that provide the information into a single output which is used by the classifier. Aitor Garcı́a Pablos et al. (2018) [23] has implemented a novel method with multi-domain and multilingual aspect- based sentiment analysis. The input of this method is a corpus of customer reviews that gives outputs as positive comments and negative comments for selected aspects. It is combined with the continuous word embedding and Maximum Entropy classifier based on the topic modeling approach. The contributions of the work are the minimal need for the supervision of the ABSAprocess in all unsupervised corpus of the customer reviews. The major task of the proposed method is to detect the aspects of opinions and polarities as in aspect-based methods. It guides the polarity modeling and topic with the desired level of biasing the parameters under distributions of the topics sampled. Sentiment analysis has been extensively studied for product and movie reviews. In sociologies, it is settled that feelings and suppositions assume a particular job in the public activityand correspond with the social associations. When facing the feelings, individuals do not remain quiet about the feelings, but instead, they will in general show them. Similarly, individuals tend to ”catch” others feelings as an outcome of facial, vocal, and postural input, which has been perceived as passionate infection in sociologies. Liu (2010) [11] in sentiment analysis discussed the existing methods and how the sentiment analysis can be performed for the unstructured data present in the web. Most of the sentiment analysis methods focussed only on subjective statements and give the least priority to objective statements which too can contain certain sentiments. To overcome this, a new approach was proposed that can handle both subjective and objective statements for analysis. A multiple instance learning network with a novel abstract-based memory mechanism (MILAM) was proposed to address the challenging task. It was found that the accuracy of the overall recommendation prediction achieved by their model can be improved if the borderline reviews are removed. The fundamental problem of sentiment analysis, namely sentiment polarity categorization was overcome by Xing Fang and Justin Zhan (2015). By using machine learning techniques, the sentimental analysis for social media was done in the early period. Sentence level sentiment analysis was used for their research. The system has experimented on data set of movie reviews and tweets extracted from online newspapers and Twitter. (Rani and Kumar 2018). Mumtaz and Ahuja (2014) proposed a sentimental analysis technique of movie reviews obtained from Twitter using the Senti-lexicon algorithm. The main aim of opinion mining is classifying, recognizing and obtaining opinion polarity of the given data. The earlier research works on sentiment analysis is mainly done at the document and sentence level. The common machine learning algorithms for sentiment classification is based on supervised learning models that are trained from labeled corpora where each document has been labeled as positive or negative be- fore training. It is not easy to obtain such labeled corpora in practical for sentiment analysis. Generally, the sentiment classification models trained on one domain will not produce better results when shifted to other domains.

C. Hybrid Methods

Mcauliffe and Blei (2007) [24] suggested the Supervised Latent Dirichlet Allocation (SLDA), a statistical model that works with labeled documents. The model takes into account, different types of response which motivates the use of the maximum-likelihood procedure for estimation of parameters. The proposed work depends on the variational approximations to handle inflexible latter expectations. Devi Parikh (2007) [25] focused on the ensemble of clas- sifier algorithms, especially for incremental learning. The proposed model uses machine learning methods as base clas- sifiers. Many statistical tests were conducted to determine optimum and moderate design parameters for error goal. It also determines the number of hidden layer nodes and the number of classifiers required for each feature set. The algorithm learns from data iteratively in a sequential manner. It includes the dataset with heterogeneous features and generates ensemble classifiers for each dataset through a modified weighted majority voting method. Ying–Lang Chang et al. (2008) [26] proposed a new topic model using LDA which is adaptive to meet the expectations from the unknown domain knowledge in real-world applications. Reimbursing such a domain mismatch assures a good document representation at different time stamps. It uses a recursive Bayes algorithm to implement an Adaptive Topic Model (ATM). LDA is applied to new domains with these adaptive models where the data is matched to topic and domain from time to time without waiting for long batch data. By properly characterizing LDA parameters with appropriate priors, the reproducible posterior distributions are derived for efficient implementation of ATM. In the proposed work, ATM continuously captured the evolution of the topic and consistently improved the document modeling and categorization from new domain data . Daume et al. (2010) [27] proposed a semi-supervised al- gorithm that uses labelled data in source and both labeled and unlabeled data in the target domain. It is an extension of a well-known supervised domain adaptation approach.

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

The introduced domain adaptation approach is simple to implement and can be applied as a pre-processing step to any supervised learner. Most of the research works imply that cross-domain sentiment classification has recently received more attention with the domain adaptation than single domain classifiers in terms of classification accuracy.

Dingcheng Li et al. (2011) [18] proposed a novel application of topic models to perform the Entity Relation Detection (ERD). It is introduced to make use of the latent semantics of text. The task of relation detection is viewed as a topic modeling problem in the new method. The underlying topics of the document are identified as indications of relations between named entities. In the proposed the pairs of named entities and features associated with them are considered as mini documents. Lin et al. (2012) [28] proposed a new probabilistic modeling framework Joint Sentiment-Topic (JST) model based on LDA. It detects topic and topic sentiment together from text document which has opinions. Reverse-JST, the modified version of the JST model is obtained by reversing the order of sentiment and topic generation in the modeling process is also proposed. Many experiments revealed that when sentiment priors are added to the JST model it shows comparable performance than Reverse JST. The method also provides a statement that weakly supervised JST produces satisfactory results when trained in one domain and shifted to others compared to other supervised approaches. The results are verified by testing the model with data sets from five different domains. Zhu et al. (2012) [29] proposed the Maximum Entropy Discrimination Latent Dirichlet Allocation (MEDLDA) which is a supervised topic model. It uses the maximum margin principle of SVM and the hierarchical Bayesian model (LDA) effective topic modeling. It attempts to use side information during the estimation of latent topical representations. MedLDA is designed to use the discriminative max-margin learning technique in a probabilistic framework unlike existing supervised topic models. Commonly used topic models first detect a latent topic for each document using a topic model and then input the identified topic to another downstream prediction model, but in this model, it is done in a single framework. Maite Taboada (2011) [10] presents a lexicon-based approach to extract sentiment from text and the Semantic Orientation Calculator (SO CAL) is used to mark the words in dictionaries to increase their orientation. The SO-CAL is applied in the polarity classification of determining the positive or negative aspects assigned in the text. The author comes up with the recommendation of an extra substantive discourse of utilizing a nonexclusive positive word. It suggests the identification maximum farthest point on the separation of redundancies in the sentiment context and decreases the weight of the words that express the same meaning to a topic. Monotonous weighting does not have any significant attitude to the irrelevant words, but the method is trained to draw extraordinary consideration on such words. Another motivation is to reduce the words that show up frequently in content and have neutral meaning towards the context.

Ghiassi (2018) [30] presented a Twitter Sentiment analysis using a Supervised Machine Learning Approach in a transferable Lexicon set. It recognizes the three basic limitations and address these issues in the proposed work: (1) accomplishing a high level of exactness as estimated by standard measurements (2) meeting an additional metric standard that guarantees the closeness of at least one critical element in each tweet, and (3) fulfilling the ”space transferability” or inclusion of all states of the list with of crosswise capabilities over different areas. To accomplish these limitations, a new procedure is presented that parts the choice of highlights into a reusable list of capabilities (Twitter Generic Feature Set, TGFS) that can be used crosswise over different areas. The Twitter explicit component designing procedure presented in this exploration applies the conventional computational semantic measures and progressively diminishes the list of capabilities by presenting ”highlight gathering” and ”meta-highlights” to speak to a lot of n-grams. Bryan Li (2019) [31] has present Acoustic and Lexical Sentiment Analysis for Customer Service Calls for the increasing development of the sentiment analysis for customer service. The acoustic model is anxious about relating highlights of something that is talked about. This framework utilizes every expression as a smaller one than a normal group of casings. People utilize both acoustic and lexical models to pick on educated choices about the conclusion regarding articulations. Atanu (2018) [32] presents a Senti-N-Gram in lexicon for the Sentiment analysis. It is used for the sentiment analysis of the products services based on trained datasets. The proposed system as has two sections: (1) a strategy for n-gram sentiment dictionary creation or Senti-N-Gram development and (2) a system for opinion characterization. The initial segment proposes n-grams development and their score extraction strategy by separating the purchaser audits and related evaluations. The second part proposes another way to deal with the characterisation of opinions in the surveys. Compared to unigrams with invalid sentiments for some instances, bigrams proved to be suitable for effective analysis.

D. Ensemble Methods

Ho (2002) introduced the term “ensemble methods”, which usually refers to the collections of classifiers. These classifiers are minor variations of the same classifiers. In multiple classifier systems, it is viewed as a broader category that includes the hybridization of different models. The proposed method focuses on variation-based ensemble methods concerning the data. It manipulates the training examples in such a way that each classifier is trained with a different training set.

AdaBoost and Bagging are the most commonly used ensemble learning algorithms, but there exist many variants and other different approaches. Devi Parikh (2007) [25] focused on the ensemble of classifier algorithms, especially for incremental learning. The proposed model uses machine learning methods as base classifiers. Many statistical tests were conducted to determine optimum and moderate design parameters for error goal. It also determines the number of hidden layer nodes and the number of classifiers required for each feature set. The algorithm learns from data iteratively in a sequential manner. It includes the dataset with heterogeneous features and generates ensembleclassifiers for each dataset through a modified weighted majority voting method. Raimon Bosch et al. (2013) [33] proposed a model to deal with analysing short messages about brands in twitter and to classify them as positive and negative using Senti WordNet. After several experiments, it is suggested to use a semi-supervised approach which could increase the quality of the dictionary and it could be adapted to any specific domain. As the existing methods do not have strong grammatical structures inside tweets, an approach based on structured N-grams was proposed. A new model called sentigram is proposed, which is the aggregation of several N-grams. The proposed approach allows developing models that are very precise to specific domains and also captures the relation between aspects and sentiment words. Cagatay Catal and Mehmet Nangir (2016) [34] have intro- duced as novel classification techniques with multiple classifier systems concept on the Turkish sentiment classification problem. Majority vote algorithm in ensemble method is used in combination with three classifiers, namely Naive Bayes, Support Vector Machine (SVM), and Bagging. Parameters of the SVM are optimized when it is used as an individual classifier. The examined results of the model showed that multiple classifier systems can increase the performance of individual classifiers on Turkish sentiment classification datasets. The proposed approach achieved better performance than individual classifiers (Naïve Bayes and SVM) on these datasets. Lei Zhang (2018) [11] presented a deep learning technique for sentiment analysis or decision mining. It is the computational examination of people's assessments, assumptions, feelings, evaluations, and the behaviours towards substances, for example, items, administrations, affiliations, people, issues, occasions, themes, and their characteristics. The initiation and quick development of the field concur with those of the internet-based life on the Web, such as audits, gathering arguments blogs, smaller-scale sites, Twitter, and interpersonal organizations, without precedent for mankind's history. It is generally considered in data mining, web mining, content mining, and data recovery applications. Koyel ChakraBorty (2018) [35] provides an insight of recent features as Machine learning culture in the set of movie reviews using the deep learning methods. It is a profound learning developed to overcome the drawbacks of machine learning methods. The proposed systems can be imagined as a strategy implemented to find highlights utilizing a few phases of uneven tasks in the highest dimension which starts from the lower level. The work employs the use of Word2vec model of word representation. Word2vec is a case of a computationally able diagnostic model in which word insertion is gained from the unstructured content. The principle advantage for learning highlights in Word2vec model is that there is no necessity of a completely probabilistic model. It introduces two different representations: Continuous Bag-of-Words model (CBOW) and the Skip-Gram model. Oscar Arague (2017) [36] presented a enhancing deep learning methods using in sentiment analysis for the social applications. The authors proposed a combination of two primary assumption examination approaches through a few outfit models in which the data given by numerous collections of highlights. Twitter and movie reviews datasets are examined on the proposed model and it gave a quantifiable report about the results of these combined models. Ensemble methods improve machine learning results by combining multiple models. The ensemble of models gives better performance as compared to the individual models in most of the cases. Computation and design time of ensemble method is generally high. Choosing the models plays a crucial role in improvement of classification in sentiment analysis.

S.No.	Title [Reference]	Year	Approach	Result
1.	Web Mining and Minimization Framework Design on Sentimental Analysis for Social Tweets Using Machine Learning [53]	2019	A sentimental tweets segregation and classification based on content I proposed under the objective of web minimization for optimized search results	97.82% accuracy via real-time tweeter logs
2.	An optimal support vector machine based classification model for sentimental analysis of online product reviews [56]	2020	Support Vector Machine (SVM) based classification model is ap- plied to classify the product re- views and K-means clustering tech- nique is applied to cluster the avail- able data into two groups	The simulation outcome pointed out the superior characteristics of the pre- sented model under sev- eral aspects

3.	Sentimental Short Sentences Classification by Using CNN Deep Learning Model with Fine Tuned Word2Vec [57]	2020	Word2Vec word embedding and Convolutional Neural Network (CNN) method are be implemented for effective text classification	The trained model is providing 82.19% accuracy for testing samples
4.	Sentimental analysis of Indian regional languages on social media [55]	2021	Provides the sentimental analysis of regional languages on Twitter data. Initially, the customer reviews are scraped from Twitter, and by using suitable algorithms and natural language processing techniques the reviews are classified based on the sentiment behind the statement.	Achieved 98% accuracy
5.	Explicit quantification of coastal cultural ecosystem services: A novel approach based on the content and sentimental analysis of social media [54]	2022	Applying natural language processing to apps with machine learning in cloud computing, authors derived the sentiments based on user-generated textual content associated with coastal ecotourism	Results show that hotspots of cultural ecosystem services (CESs) are spatially concentrated in both cultural attractions and protected areas, which are critical for coastal ecosystem management and protection

E. Other Previous

Lexicon-based methods require predefined sentiment lexicon to determine the polarity of any document. However, the accuracy of lexicon-based method is reduced drastically in the presence of emoticons and short hand texts, as they are not the part of predefined sentiment lexicon [37]. Emoticons are the visual emotional symbols used by the users at social media. Hu et al. [38] proposed a novel method of sentiment analysis that considers the short texts like “gudnite” and emotional symbols such as “:)”, in a unified frame-work. This problem can be resolved by examining the contributions of other emotion indication information existing in social media, like product ratings, restaurant reviews, and other emotion correlation information [39] [40] such as correlation between two words in a post. More-over, emotion correlation for posts are usually represented by a graph in which nodes represent the data points and edge represent correlation between the words. Canuto et al. [41] proposed a new sentiment- based meta-level features for effective sentiment analysis. This method has a capability to utilize the information from the neighborhood effectively and efficiently to capture important information from highly noise data. Kontopoulos et al. [42] proposed ontology-based sentiment analysis of tweets. In this method, a sentiment grade has been assigned for every distinct notion in the tweets. Further, Mohammad et al. [43] analyzed US presidential electoral tweets by using supervised automatic classifiers and identified the emotional state, emotion stimulus, and intent of these tweets. Coletta et al. [44] combined the strength of SVM classifier with a cluster ensemble for refining the tweet classification. SVM classifier is executed first to classify tweets, thereafter C3E-SL algorithm has been used to enhance the classification of tweets. Agarwal et al. [45] introduced a new sentiment analysis model based on common-sense information mined from ConceptNet-based ontology and context knowledge. ConceptNet-based ontology is used to discover the domain specific concepts which is further used to obtain the domain specific important features. Saif et al. [46] proposed a SentiCircle method which assigns context-specific sentiment orientation to words. SentiCircle method has been introduced to update the sentiment strength of many terms dynamically. Fernandez et al. [47] introduced a novel unsupervised method based on linguistic sentiment propagation model to predict the sentiments in informal texts. Due to unsupervised nature, this method does not require any training and uses linguistic content for sentiment analysis. Previous research efforts on this area includes, on the one hand, Hogenboom et al. [48] focuses in using rhetorical structure in sentiment analysis, and utilises structural aspects of text as an aid to distinguish important segments from those less important, as far as contributing to the overall sentiment being communicated. As such, they put forward a hypothesis based on segments’ rhetorical roles while accounting for the full hierarchical rhetorical structure in which these roles are defined. Heerschop et al. [49] propose a Rhetorical Structure Theory (RST) based approach, called Pathos, to perform document sentiment analysis partly based on the discourse structure of a document.

Text is then classified into important and less important spans, and by weighting the sentiment conveyed by distinct text spans in accordance with their importance, the authors claim that they can improve the performance of a sentiment classifier. The work by Bravo-Mffarquez et al. [50], on the use of multiple techniques and tools in SA, offers a complete study on how several resources that are focused on different sentiment scopes can complement each other. The authors focus the discussion on methods and lexical resources that aid in extracting sentiment indicators from natural languages in general. Schouten et al. [51] provides a complete survey specific to aspect-level sentiment analysis. A number of researchers have explored the application of hybrid approaches by combining various techniques with the aim of achieving better results than a standard approach based on only one tool. Indeed, this has been done by Poria et al. [52] where a novel framework for concept-level sentiment analysis, Sentic Pattern, is introduced by combining linguistics, common-sense computing, and machine learning for improving the accuracy of tasks such as polarity detection. The authors claim that by allowing sentiments to flow from concept to concept based on the dependency relation of the input sentence, authors achieve a better understanding of the contextual role of each concept within the sentence and, hence, obtain a polarity detection engine that outperforms state-of-the-art statistical methods.

III. DISCUSSION

The opinion mining methods discussed considers only the opinions and not the topics which add value to the classifica- tion. The performance of classification is considerably less in all the above methods. Most of the previous researches are based on supervised learning. It requires labeled data for training and the accuracy of classification is still a challenge. They focused only on classifying opinions but do not concentrate on topics which reduce the effectiveness of classification. Most of the works used lexicon or WordNet for classification and only unigrams were considered for sentiment analysis which does not give accurate sentiment analysis for negated words. In the previous section various research works pertaining to sentiment analysis, topic detection and hybrid methods are reviewed. Various literature on classification of sentiments using various machine learning algorithms is discussed. Classical machine learning algorithms classify the sentiments based on polarity alone, which is not sufficient for the customer. Sentiments related to topics were considered to provide more useful information and in that connection topic modeling methods were discussed. The extant methods do not employ classifiers. Hence, there is a need to integrate the exist- ing machine learning classifiers with topic model. Finally, ensemble methods were discussed, which combine various classical machine learning algorithms for improving accuracy. It is believed that ensemble methods provide better results in classification when compared to single algorithm.

IV. CONCLUSION

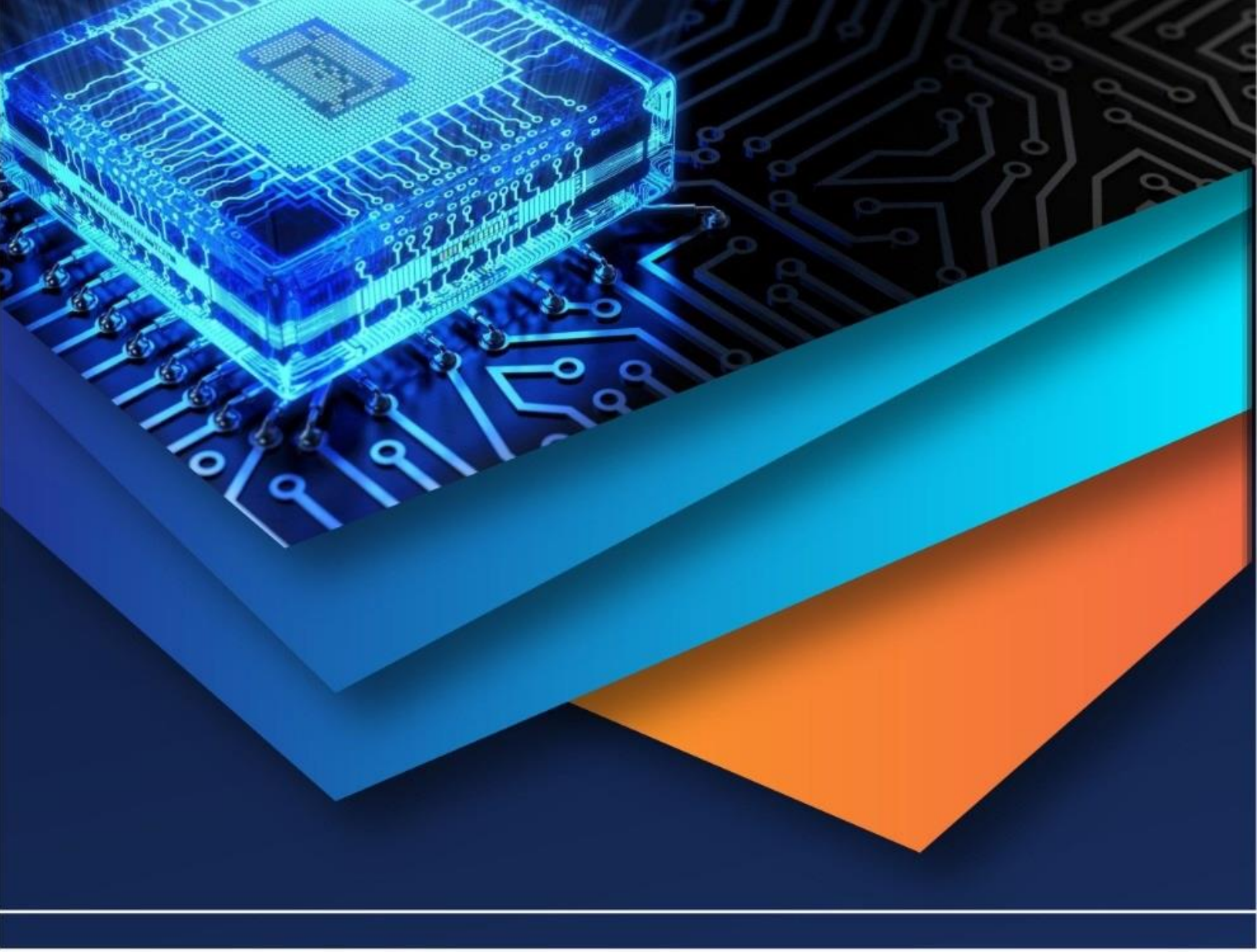
Sentiment analysis is the measurement of positive and negative language. It is a way to evaluate written or spoken language to determine if the expression is favourable, unfavourable, or neutral, and to what degree. It is one of the prominent fields of data mining that deals with the identification and analysis of sentimental contents generally available at social media. In this work, a survey has been done on various sentiment analysis approaches proposed in the past. It is concluded from the review that hybrid models are found to be better as compared to opinion based or simple machine learning models. It is also seen that use of ensemble model rather than simple machine learning models will prove better for the same.

REFERENCES

[1] P. N. Howard, “The arab springs cascading effects.” Pacific Standard, vol. 23, 2011.
[2] E. Sulis, D. Irazú Hernández Farías, P. Rosso, V. Patti, and G. Ruffo, “Figurative messages and affect in twitter: Differences between irony, sarcasm and not,” Knowledge-Based Systems, vol. 108, pp. 132–143, 2016, new Avenues in Knowledge Bases for Natural Language Processing. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0950705116301320
[3] A. Reyes, P. Rosso, and D. Buscaldi, “From humor recognition to irony detection: The figurative language of social media,” Data and Knowledge Engineering, vol. 74, pp. 1–12, 2012, applications of Natural Language to Information Systems. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0169023X12000237
[4] J. Bollen, H. Mao, and X. Zeng, “Twitter mood predicts the stock market,” Journal of Computational Science, vol. 2, no. 1, pp. 1–8, 2011. [Online] Available: https://www.sciencedirect.com/science/article/pii/S187775031100007X
[5] S. Asur and B. A. Huberman, “Predicting the future with social media,” in 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, vol. 1, 2010, pp. 492–499.
[6] A. Hasan, S. Moin, A. Karim, and S. Shamshirband, “Machine learning-based sentiment analysis for twitter accounts,” Mathematical and Computational Applications, vol. 23, no. 1, 2018. [Online]. Available: https://www.mdpi.com/2297-8747/23/1/11
[7] B. Pang, L. Lee, and S. Vaithyanathan, “Thumbs up? Sentiment classification using machine learning techniques,” in Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP 2002). Association for Computational Linguistics, Jul. 2002, pp. 79–86. [Online]. Available: https://aclanthology.org/W02-1011

[8] D. Ramage, D. Hall, R. Nallapati, and C. D. Manning, "Labeled LDA: A supervised topic model for credit attribution in multi-labeled corpora," in Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing. Singapore: Association for Computational Linguistics, Aug. 2009, pp. 248–256. [Online]. Available: https://aclanthology.org/D09-1026
[9] K. Krishnakumari, "Domain adaptation in sentiment classification using deep sentence properties," Global Journal of Pure and Applied Mathematics, 01 2017.
[10] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, "Lexicon-Based Methods for Sentiment Analysis," Computational Linguistics, vol. 37, no. 2, pp. 267–307, 06 2011.
[11] L. Zhang and B. Liu, Sentiment Analysis and Opinion Mining. Boston, MA: Springer US, 2017, pp. 1152–1161.
[12] B. Keith Norambuena, E. Lettura, and C. Villegas, "Sentiment analysis and opinion mining applied to scientific paper reviews," Intelligent Data Analysis, vol. 23, pp. 191–214, 02 2019.
[13] R. Bajpai, D. Hazarika, K. Singh, S. Gorantla, E. Cambria, and R. Zimmerman, "Aspect-sentiment embeddings for company profiling and employee opinion mining," 02 2019.
[14] O. Joshi and G. Simon, "Sentiment analysis tool on cloud: Software as a service model," 02 2018, pp. 459–462.
[15] C. Lin, "Joint sentiment/topic model for sentiment analysis," International Conference on Information and Knowledge Management, Proceedings, 01 2009.
[16] S. Rüping, "Incremental learning with support vector machines," 11 2001.
[17] F. Fdez-Riverola, E. Iglesias, F. Díaz, J. Méndez Reboredo, and J. Corchado Rodríguez, "Applying lazy learning algorithms to tackle concept drift in spam filtering," Expert Syst. Appl., vol. 33, pp. 36–48, 07 2007.
[18] D. Li, S. Somasundaran, and A. CHAKRABORTY, "Erd-medlda: Entity relation detection using supervised topic models with maximum margin learning," Natural Language Engineering, vol. 18, 04 2012.
[19] J. Liang, F.-y. Zhang, X. Xiong, X. Chen, L. Chen, and G.-h. Lan, "Manifold regularized proximal support vector machine via generalized eigenvalue, International Journal of Computational Intelligence Systems, vol. 9, pp. 1041–1054, 11 2016.
[20] N. Shafaf and H. Malek, "Applications of machine learning approaches in emergency medicine; a review article," Archives of academic emergency medicine, vol. 7, p. 34, 06 2019.
[21] S. Kotsiantis, "Increasing the accuracy of incremental naive bayes classifier using instance based learning," International Journal of Control, Automation and Systems, vol. 11, 02 2013.
[22] Y. Ma, H. Peng, T. Khan, E. Cambria, and A. Hussain, "Sentic lstm: a hybrid network for targeted aspect-based sentiment analysis," Cognitive Computation, vol. 10, pp. 1–12, 08 2018.
[23] A. García-Pablos, M. Cuadros, and G. Rigau, "W2vlda: Almost unsupervised system for aspect based sentiment analysis," Expert Systems with Applications, vol. 91, 05 2017.
[24] D. Blei, A. Ng, and M. Jordan, "Latent dirichlet allocation," vol. 3, 01 2001, pp. 601–608.
[25] D. Parikh and R. Polikar, "An ensemble-based incremental learning approach to data fusion," IEEE transactions on systems, man, and cybernetics. Part B, Cybernetics : a publication of the IEEE Systems Man, and Cybernetics Society, vol. 37, pp. 437–50, 05 2007.
[26] H. Jelodar, Y. Wang, C. Yuan, and X. Feng, "Latent dirichlet allocation (lda) and topic modeling: models, applications, a survey," 11 2017.
[27] H. Daumé III, A. Kumar, and A. Saha, "Frustratingly easy semi-supervised domain adaptation," in Proceedings of the 2010 Workshop on Domain Adaptation for Natural Language Processing. Uppsala, Sweden: Association for Computational Linguistics, Jul. 2010, pp. 53–59. [Online]. Available: https://aclanthology.org/W10-2608
[28] C. Lin, R. Everson, and S. Rueger, "Weakly supervised joint sentiment-topic detection from text," IEEE Transactions on Knowledge and Data Engineering, vol. 24, pp. 1134–1145, 01 2011.
[29] J. Zhu, A. Ahmed, and E. Xing, "Medlda: maximum margin supervised topic models for regression and classification," 01 2009, p. 158.
[30] M. Ghiassi and S. Lee, "A domain transferable lexicon set for twitter sentiment analysis using a supervised machine learning approach," Expert Systems with Applications, vol. 106, pp. 197–216, 2018.
[31] B. Li, D. Dimitriadis, and A. Stolcke, "Acoustic and lexical sentiment analysis for customer service calls," 05 2019, pp. 5876–5880.
[32] A. Dey, M. Jenamani, and J. Thakkar, "Senti-n-gram: An n-gram lexicon for sentiment analysis," Expert Systems with Applications, vol. 103, 03 2018.
[33] R. Bosch, "Sentiment analysis : Incremental learning to build domain models," 2014.
[34] C. Catal and M. Nangir, "A sentiment classification model based on multiple classifiers," Applied Soft Computing, vol. 50, pp. 135–141, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1568494616305919
[35] K. Chakraborty, S. Bhattacharyya, R. Bag, and A. E. Hassanien, Comparative Sentiment Analysis on a Set of Movie Reviews Using Deep Learning Approach, 01 2018, pp. 311–318.
[36] O. Araque, I. Corcuera-Platas, J. F. Sánchez-Rada, and C. A. Iglesias, "Enhancing deep learning sentiment analysis with ensemble techniques in social applications," Expert Systems with Applications, vol. 77, pp. 236–246, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417417300751
[37] L. Zhang, R. Ghosh, M. Dekhil, M. Hsu, and B. Liu, "Combining lexicon-based and learning-based methods for twitter sentiment analysis," 01 2011.
[38] T.-J. G. H. Hu, X. and H. Liu, "Unsupervised sentiment analysis with emotional signals." in In Proceedings of the 22nd international conference on world wide web, 2013, p. 607–618.
[39] T.-L. T. J. Hu, X. and H. Liu, "Exploiting social relations for sentiment analysis in microblogging." in In Proceedings of the sixth ACM international conference on web search and data mining, 2013, p. 537–546.
[40] N. N. Yusof, A. Mohamed, and S. Abdul-Rahman, "Reviewing classification approaches in sentiment analysis," in Soft Computing in Data Science. Singapore: Springer Singapore, 2015, pp. 43–53.
[41] G.-M. A. Canuto, S. and F. Benevenuto, "Exploiting new sentiment-based meta-level features for effective sentiment analysis." in In Proceedings of the ninth ACM international conference on web search and data mining, 2016, p. 53–62.
[42] E. Kontopoulos, C. Berberidis, T. Dergiades, and N. Bassiliades, "Ontology-based sentiment analysis of twitter posts," Expert Systems with Applications, vol. 40, no. 10, pp. 4065–4074, 2013.

[43] S. M. Mohammad, X. Zhu, S. Kiritchenko, and J. Martin, "Sentiment, emotion, purpose, and style in electoral tweets," Information Processing and Management, vol. 51, no. 4, pp. 480–499, 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306457314000880
[44] L. F. S. Coletta, N. F. F. da Silva, E. R. Hruschka, and E. R. Hruschka, "Combining classification and clustering for tweet sentiment analysis," in 2014 Brazilian Conference on Intelligent Systems, 2014, pp. 210–215.
[45] M. N. B. P. Agarwal, B. and S. Garg, "Sentiment analysis using common-sense and context information." Computational Intelligence and Neuro- science, vol. 30, 2015.
[46] H. Saif, Y. He, M. Fernandez, and H. Alani, "Contextual semantics for sentiment analysis of twitter," Information Processing and Management, vol. 52, no. 1, pp. 5–19, 2016, emotion and Sentiment in Social and Expressive Media. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306457315000242
[47] M. Fernández-Gavilanes, T. Álvarez López, J. Juncal-Martínez, E. Costa-Montenegro, and F. Javier González-Castaño, "Unsupervised method for sentiment analysis in online texts," Expert Systems with Applications, vol. 58, pp. 57–75, 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417416301300
[48] A. Hogenboom, F. Frasincar, F. de Jong, and U. Kaymak, "Using rhetorical structure in sentiment analysis," Communications of the ACM, vol. 58, pp. 69–77, 06 2015.
[49] B. Heerschop, F. Goossen, A. Hogenboom, F. Frasincar, U. Kaymak, and F. de Jong, "Polarity analysis of texts using discourse structure," 10 2011, pp. 1061–1070.
[50] F. Bravo-Marquez, M. Mendoza, and B. Poblete, "Meta-level sentiment models for big social data analysis," Knowledge- Based Systems, vol. 69, pp. 86–99, 2014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0950705114002068
[51] K. Schouten and F. Frasincar, "Survey on aspect-level sentiment analysis," IEEE Transactions on Knowledge and Data Engineering, vol. 28, no. 3, pp. 813–830, 2016.
[52] S. Poria, E. Cambria, G. Winterstein, and G.B. Huang, "Sentic patterns: Dependency-based rules for concept-level sentiment analysis," Knowledge-Based Systems, vol. 69, pp. 45–63, 2014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S095070511400183X
[53] T. S. Raghavendra and K. G. Mohan, "Web mining and minimization framework design on sentimental analysis for social tweets using machine learning," Procedia Computer Science, vol. 152, pp. 230–235, 2019, international Conference on Pervasive Computing Advances and Applications- PerCAA 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050919306994
[54] H. Cao, M. Wang, S. Su, and M. Kang, "Explicit quantification of coastal cultural ecosystem services: A novel approach based on the content and sentimental analysis of social media," Ecological Indicators, vol. 137, p. 108756, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1470160X22002278
[55] K. Rakshitha, R. HM, M. Pavithra, A. HD, and M. Hegde, "Sentimental analysis of indian regional languages on social media," Global Transitions Proceedings, vol. 2, no. 2, pp. 414–420, 2021, international Conference on Computing System and its Applications (ICCSA-2021). [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2666285X21000674
[56] P. Vijayaragavan, R. Ponnusamy, and M. Aramudhan, "An optimal support vector machine based classification model for sentimental analysis of onlin product reviews," Future Generation Computer Systems, vol. 111, pp. 234–240, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167739X19333138
[57] A. K. Sharma, S. Chaurasia, and D. K. Srivastava, "Sentimental short sentences classification by using cnn deep learning model with fine tuned word2vec," Procedia Computer Science, vol. 167, pp. 1139–1147, 2020, international Conference on Computational Intelligence and Data Science. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050920308826



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)