

Sentiment Analysis Approaches and its Recent Advancements: A Review of Research Paradigm

Sherin A, Dr. I. Jasmine Selvakumari Jeya

* (Department of Information Technology, Hindusthan College of Engineering and Technology, Coimbatore
Email: wjasminejeya@gmail.com)

** (Department of Computer Science and Engineering, Hindusthan College of Engineering and Technology, Coimbatore
Email: sherinameer@gmail.com)

Abstract:

The recent advancement in Sentiment Analysis (SA) enables to discern the textual information for identifying the sentiments or opinions from the text. It also helps to extract subjective information from the text. The proliferation of online communication medias like Twitter, Facebook, YouTube, and so forth provides a platform for user's views, thoughts, emotions, and opinions and are these days a lot of impact into human life. Nowadays Sentiment Analysis emerged as an active research area for analysing and making sense of these user generated data and thereby providing valuable insights for decision making capabilities for peoples and organisations. The proposed review paper, tend to match on opinion mining or feeling assessment which is an area of Natural language Processing and Machine Learning. Analysing sentiments articulated in the user generated content is beneficial and challenging yet practically cumbersome. Here a comprehensive study of various Sentiment Analysis methodologies, their recent advancements and open problems are performed to play out an evaluation study.

Keywords —Sentiment Analysis, Opinion Mining, feeling mining, Emotional Artificial Intelligence, Lexicon based approach, Machine Learning based Approaches, Support Vector Machine (SVM), Naïve Bayes (NB), Neural Networks.

I. BACKGROUND

Sentiments of people can be articulated in diverse ways. It can be stated with various types of emotions, by passing judgments, vision or insight or opinions of people [1]. An emotion can be expressed as unexpected reaction by person intentionally or unintentionally depending on circumstances. If a study of the emotion in text format is executed then it can be perceived in two different ways. Primarily, it can be studies as impact on writer – the way he picks word to express certain emotions. And secondly impact on reader-

how he infers content written depending on his state of mind and his ability to analysis things. Sentiment analysis or feeling mining is the method to identify and recognize or categorize the users 'emotion or feeling for any services like movies, product reviews, events or any attribute is positive, negative or neutral. The main sources of data are from sites like Facebook, Twitter, blogs, microblogs, reviews, forum discussions and so on. It can be even applicable to stock markets [2,3], news articles [4] and Political debates as well. In political debates for example, we could figure out people's opinions on a particular election candidates/representatives or

political parties. This innovative research field is very prevalent currently because of its opinionated data where user can discover reviews for any services which are useful for their day-to-day activities. The large amount of opinionated data is stockpiled in digital forms. For specific topic or any opinion, the sentiment analysis which relates the mining of the data works and give the output. For sentiment analysis, research exploration works on emotion-based summarization, feeling or mind extraction. Sentiment analysis also known as Opinion mining or Emotion mining or Feeling mining which uses NLP – Natural Language Processing or state of the art Machine learning to following the emotions, feelings of the public opinion about a specific area for any product or services.

There are many real time applications for Sentiment Analysis. It can be effectively used in Decision Making Support, Business, Prediction and trend analysis.

A. Decision making support

Building a website that could perform decision making is actually a crucial part in sentiment analysis application. Analysis should be performed in such a way that it has its own advantage like; it can lead to different ideas which can help users to make genuine decision in their daily life such as choosing a good hotel to go for dinner, or buying a new bike or choosing a good movie to watch and so on.

B. Business related application

Because of every day fluctuating market, the competition in business domain has increased a lot in co-operative world. Every company wants to create an innovative product which can fully satisfy their customers' expectations. To attain more valuation of their product, organization can accumulate all the needs of their users and enhance the efficiency of product from customer feedback

collected from their customers or through recommender systems.

C. Predictions and trend analysis

Anticipating trend in the market and tracking views of public by sentiment scrutiny is very challenging in current situation

II. SENTIMENT ANALYSIS

Sentiment Analysis (also known as opinion mining or Emotion Artificial Intelligence) is an evolving and a dynamic research ground which encompasses the usage of natural language processing, text analysis, computational linguistics, and machine learning languages to scientifically identifying the emotional states and subjective information from the texts [5]. As Sentiment analysis is an eminent research area, it can be also beneficial in numerous ways in surveys and advertisement campaign by receiving the success rate of any product or services with people 's opinion or suggestion. It also gives the information about people liking and disliking so that company come to be much clear and precise idea regarding its product features. Sentiment Analysis has augmented a lot of acceptance among various zone like politics [6], business [7] and marketing/selling and advertisement (to estimate sales of specific products). So, identifying type of sentence is the most significant part of Sentiment Analysis. So, the initial Procedure in Sentiment Analysis is to classify the sentence either subjective or objective [8]. Current research is advanced both on supervised and unsupervised learning technique to provide polarity scores of sentiments.

In preliminary research all or amalgamation of supervised as well as unsupervised techniques are used like Support vector machine, Maximum Entropy, Naive Bayes, Exploit sentiment lexicons, Grammatical analysis and syntactic patterns. In general, —Sentiment Analysis comprises advanced processes and has a totally different series of tasks, sentiment classification (supervised or unsupervised), subjective or objective analysis, and opinion of feeling extraction. For any text

document or a sentence, to do the labelling as subjective or objective can be assessed by subject level analysis. All the sentences are classified into neutral, negative or positive sentiments based on the emotions from the user opinions or reviews [9].

Subjectivity detection is the task of classifying objective and subjective sentences. Objective sentences are those judgements which do not exhibit any sentiment, and thus omitted while performing the analysis. So, it is inevitable for a sentiment analysis engine to discover and separate the objective sentences for further analysis and study, e.g., polarity detection. In subjective sentences, opinions can frequently be expressed on one or multiple topics. The main purpose of Sentiment Analysis is to split the entire documents or sentences into either of the two classes; objective or subjective. Normally in Sentiment Analysis or in Text classification, the sentences in the text are labelled as either subjective or objective. Sentiment analysis involves the extraction of subjective information and removal of factual information from the text. When it contains emotions or feelings subjective sentences are imperative to the procedure with sentiment analysis. Subjective sentence is made up with following views of users, viewpoints, judgements, comments and thoughts about the sentence level. Consider the following example. 1) Subjective: Harry Potter is an awesome movie. (This sentence has an emotion (awesome); thus, it is subjective). 2) Objective: Daniel Radcliffe is the actor of Harry Potter. (Here there is no sentiment hence, it is an actual fact, and thus objective) [10]

Sentiment analysis has been explored primarily at three classification levels of Granularity: 1) Document level 2) the Sentence level 3) Attribute or Aspect or Entity Level. Document level Sentiment Analysis involves the polarity identification for a particular document. Throughout this level, the whole document is taken into consideration and figure out the polarity and classifies whether the available opinion or emotions is a positive sentiment, neutral or negative

sentiment. For evaluating these type of sentiment Analysis, the whole document should be on a single subject. Sentence Level Sentiment Analysis involves the determination of the polarity of each sentence and figure out the polarity of sentence and classifies whether the available opinion or emotions is a positive sentiment, neutral or negative sentiment. Nevertheless, there is no fundamental variance between document level and sentence level classifications because sentences are in fact short documents in general [11]. Aspect or Feature or Entity Level Sentiment Analysis Involves opinion mining and summarization based on feature. The classification concerns by identifying and mining product features from the source data. This type of analysis is used when there is a need for sentiments about desired aspect or feature or entity in a target data like reviews, feedbacks, comments and complaints.

III. SENTIMENT ANALYSIS: TRENDS, METHODOLOGIES AND CURRENT PRACTICES

Numerous methodologies are offered for Sentiment Analysis, but machine learning approach (Automated) and lexicon-based method (linguistically-inclined) are most widely used for Sentiment Analysis [12] (fig 1). From the subsequent techniques, the features of text or sentences can be extracted.

A. Lexicon Based Approach

This technique computes the sentiment orientations or polarity scores of the whole document or set of sentences from semantic orientation of lexicons [13]. The dictionary of lexicons can be done manually or it can be machine generated. The WordNet dictionary is very popular and is used by many researchers [14]. First of all, lexicons are found from the whole document and then WordNet or any other kind of online thesaurus can be used to find the synonyms and antonyms to expand that dictionary. A fine-grained Sentiment Analysis of Tweets using Lexicon and machine learning

approaches are suggested in [15], in which main focus was given on dictionary-based approach. The machine learning algorithms like Support Vector Machine (SVM) and Random Forest (RF) classifiers are extensively used for sentiment classification of the Tweets. For vectorization of Tweets, Bag of Words (BoW), Term-Frequency vs. Inverse Document Frequency (TF-IDF), and a innovative feature matrices are used. In fine grained lexicon-based analysis an accuracy of 84.8% is reported. In machine learning approaches, the Support Vector Machine (kernel = linear), Support Vector Machine (kernel = RBF) and Random Forest (RF) with the Sentiwordnet feature vector got an accuracy of 92.6%, 92.9%, and 93.4%, correspondingly. Lexicon Based Sentiment Analysis using Lexicon and Latent Dirichlet Allocation (LDA) of Online Religious Lectures were performed in [16] and the main intention is to analyze sentiment on Youtube video reviews. Based on the experimental results, an accuracy of 70% is achieved in Lexicon Based Method implementation

In Dictionary-based approach [17,18], the user collected a set of sentiment orientation of context-specific words (sentiments words) and seed list is made based out of this. Subsequently, the user start searching for phrasebooks and lexicon to find synonyms and antonyms of particular text. Once this is completed, the newly created substitutes are added into the seed list. This process continues until there are no new words are found to users. In [19], a Multilingual Lexicon-based Approach for Sentiment Analysis in Social and Cultural Information System Data was performed using digital tourism data. The algorithm implemented make use of unsupervised lexicon dictionary-based approach for sentiment Analysis. The experimental results obtained demonstrate the efficiency of multilingual lexicon-based sentiment analysis algorithm.

The main limitation in dictionary-based approach is that there has to have struggle in discovering context or domain-oriented emotion words.

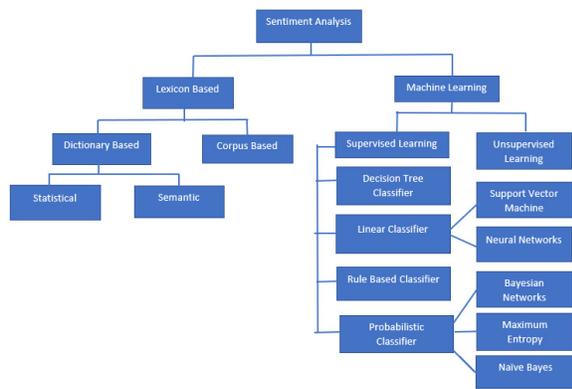


Fig 1: Classification of Sentiment Analysis Methodologies

Lexicon-based techniques use English language grammar features like adjectives and adverbs to identify the polarity scores of the text.

B) Dictionary-based approach

1)Statistical approach:The words which exhibit erratic behavior in positive behavior are assigned positive polarity. If they show negative recurrence in negative text, they have assigned with negative polarity. Sentiment analysis using cosine similarity measure proposed in [20] analysis the user sentiments in the Telecom domain which is extremely noisy, containing numerous spelling and grammatical errors, acronyms, abbreviations, shortened or slang words etc. Hence, a lexicon based pre-processing algorithm based out of statistics is proposed for noise reduction. The performance of the proposed approach is compared with some of the renowned machine learning algorithms like, Naive Bayes, Maximum Entropy and SVM. The proposed classifier out performs the above methods with an accuracy of 82.09% for the 2-class problem of identifying positive and negative sentiments.

2)Semantic approach: This approach allocates sentiment values to words and the words which are semantically closer to those words; this can be done by finding synonyms and antonyms with respect to that particular word. In [21], a semantic conceptualization method using tagged bag-of-concepts for Sentiment Analysis is proposed to identify the polarity scores of intangibly conveyed very short texts to increase the accuracy of Sentiment Analysis.

3) Corpus-based approach: Corpus is a basically a term which is a cluster of writing on a very precise matter. In this, users use the help of corpus text to draw-out the seed list which is already in organized state [22]. Sentiment analysis of online product reviews using Lexical Semantic Corpus-Based technique was proposed in [23] a web-based system which uses MySQL as a database and Python as a programming language for customers and sellers to get an analysis of the online product reviews. The client and retailer will get the visualization for the positive, neutral, and negative distributions of the product reviews.

C) Machine Learning approaches

In Machine Learning approach, tools that analyses texts for polarity, from positive to negative is used to detect sentiments automatically without human intervention. To put it simply, machine learning allows computers to learn new tasks by themselves without being expressly programmed to perform.

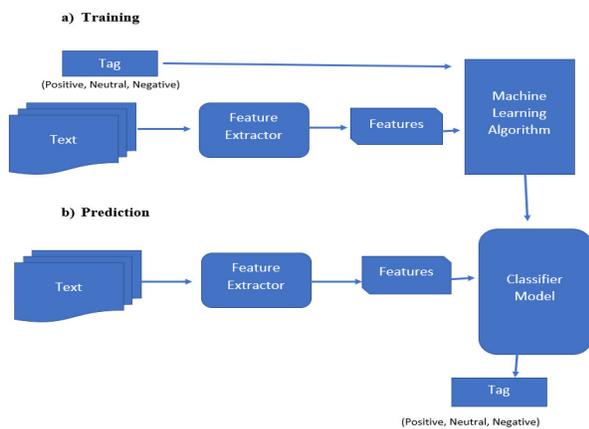


Fig 2: How Does Sentiment Analysis Work?

The most challenging open problems involved in Sentiment analysis are sarcasm, and misapplied words. For example:

“Fantastic user-friendly interface. Yeah right. An engineering degree would be helpful to improve.”

Out of context, the words ‘Fantastic user-friendly’ and ‘helpful’ could be read as positive, but this is undoubtedly a negative comment.

In Machine Learning approach, initially classification is performed by taking two different assemblies of the document. Trained data and test data are part of these as shown in Fig 2. This is termed as involuntary classification. Further text is extracted from the features using feature extraction methods and categorized into supervised and unsupervised datasets.

1)Supervised System:Among various kind of datasets, Labelled training dataset is one of them which is used in supervised system. Each type of class has its own unique property and advantages and has its label associated to its which can be used for this system. Each word, upon received is categorized under a label depending on its type and unique characteristics related with it.

1.1)Probabilistic classifier:These classifiers predict or anticipates probability function related to input records among different modules.

1.1.1)Naïve Bayes:The Naive Bayes is a classical model, which is simple and powerful technique in machine learning. The foundation for naive Bayes classification is the use of Bayes’ theorem with a strong independence assumption between the features. When used for sentiment data analysis, such as in computational linguistics, the Naive Bayes classification produces excellent results output. As Naive Bayes, a probabilistic algorithmic model that is used for sentiment analysis classification, it allocates a probability that a given word or phrase must be considered positive, neutral or negative. Naive Bayes uses the Bayes probability theorem for unknown class estimation. Essentially, the Bayes’ theorem states that “The probability of A, if B is factual, is equal to the probability of B, if A is factual, times the probability of A being factual, divided by the probability of B being factual”.

$$P(A|B) = P(B|A) \times P(A)/P(B) \tag{1}$$

Here an extensive mathematical calculation was performed. Fundamentally, Naive Bayes calculates words against each other. Therefore, with machine learning models trained for word sentiment scores, the likelihood that a word, phrase, or text is positive or negative can be efficiently calculated.

When techniques like lemmatization, stopword removal, and TF-IDF (term frequency-inverse document frequency) are executed, Naive Bayes becomes more and more predictively precise.

Advanced Naïve Bayes Algorithm Design with Part-of-Speech Tagger on Sentiment Analysis is proposed on [24] uses a POS-tagging to select certain types of words and train the sentiment classifier effectively and to reduce computation complexity.

Twitter Sentiment Analysis of Movie Reviews Using Ensemble Features Based Naïve Bayes proposed on [25] uses Naïve Bayes with Ensemble Features and a comparison of Ensemble Features and Bag of Words Features is performed. The features used for the ensemble are Twitter detailed features, written features, part of speech features, and lexicon-based features, and Bag of Words. The Experiment Results demonstrated that system f-measure value using Ensemble Features shows an accuracy of 88% while Bag of Words Features shows an accuracy of 94 %

Implementation of The Naïve Bayes Algorithm with Feature Selection using Genetic Algorithm for Sentiment Review Analysis was performed on [26], and the study uses naïve Bayes algorithm and genetic algorithms as applied feature selection for the reviewing of online fashion companies. The accuracy is calculated with confusion matrix and ROC curve and the accuracy of the model improved with the use of genetic algorithm.

A sentiment polarity detection approach presented on [27] uses Multinomial Naïve Bayes with character n-gram features for significantly better sentiment polarity scores.

1.1.2) Bayesian Network: Here the goal is to learn a network structure using dependence or independence information between set of variables. The resulting network is a directed acyclic graph (DAG), through a set of joint probability distributions, where each variable of the network is a node in the graph and the arcs between the nodes denote the probability distribution that implies the level of dependency between the nodes. Bayesian Network learns dependencies between words and their corresponding sentiment classes automatically that can be effectively used for high dimensional sentiment data. This model is very expensive and hence it's hardly used.

A Dynamic Bayesian Network Approach for Analysing Topic-Sentiment Evolution is proposed on [28] to construct a Gaussian Process Dynamic Bayesian Network to model the dynamics and interactions of the sentiment of topics on social media such as Twitter. The Dynamic Bayesian Networks helps to learn relationship between data and undoubtedly produce better classification results. The experiment demonstrates a case study of analysing the sentiment dynamics of twitter related to the event Brexit and shows better accuracy rates.

The Bayesian Network model proposed in [29] aims to develop a Bayesian-based approach for quantitative public sentiment modelling, which is capable of incorporating the inherent uncertainty of interviewed dataset and also shows better classification rates. This research work mainly focuses on Dirichlet distribution, multinomial distribution, posterior distribution and Bayesian inference; and measuring public sentiment through accumulating sampled sets of sentiment probabilities with an application-based measure.

1.1.3) Maximum Entropy: Contrasting the Naïve Baye's machine learning, Maximum Entropy makes no independence assumptions about the occurrence of words. The Maximum Entropy modelling technique delivers a probability distribution that is as close to the uniform as possible given that the distribution satisfies certain constraints or criteria. Maximum Entropy techniques proposes a unique way to combine varied pieces of contextual evidence so as to estimate the probability of a certain linguistic class happening with a certain linguistic context in which job is to estimate the probability of class 'a' occurring with context 'b'".

Tibetan Sentence Sentiment Analysis Based on the Maximum Entropy Model proposed on [30] implement maximum entropy model, for better evaluation parameter and to achieve practicability to some extent.

1.2) Linear classifier: Linear classifiers classify user generated data into labels based on a linear combination of its input features. Consequently, these classifiers separate data using a line (two dimension) or plane (three dimension) or a hyperplane (multi-dimensional) dimension). They can be used to categorize data that is linearly separable and can be even modified to classify non linearly separable user generated data.

1.2.1) Support Vector Machine (SVM): The main underlying principle of SVMs is to determine linear separators in the search space which can separate the different classes in a better way. In Fig. 3 there are 3 hyperplanes A, B and C there are 2 classes marked as "x", and "o" respectively. Hyperplane A provides the finest separation between the classes, since the normal distance of any of the data points is the largest, so it denotes the maximum margin of separation of classes. The most important purpose of this particular model is to guarantee that this is the best linear separator for classification algorithms.

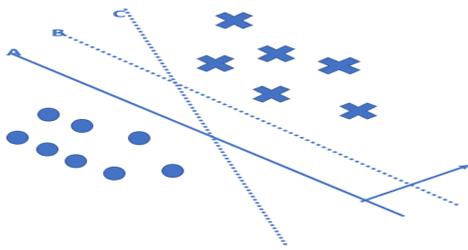


Fig 3: Using support vector machine on a classification problem.

Optimized Support Vector Machine Model for Visual Sentiment Analysis proposed in [31] developed an automatic visual sentiment analysis (VSA) model using an optimization-based support vector machine (SVM). Initially, the input images' features are extracted using weighed-FC8 layer of the pre-trained ResNet. A hybrid optimization technique called Holoentropy Life Choice Optimization (HELMCO) algorithm is used for improving the optimization. HELMCO is an ensemble of both the Life Choice Based Optimization (LCBO) algorithm and the Cross entropy (CE). The Polarity scores are measured using Emotion-6 and Abstract Art photo datasets based on performance constraints, like Accuracy, Sensitivity, and Specificity with an accuracy of 70.7% with Emotion-6 dataset and 76.8% with Art photo Dataset respectively.

Optimization Of Parameter Support Vector Machine (SVM) using Genetic Algorithm to Review Go-Jek's Services proposed in [32] uses SVM method to classify opinion data based on assessment attributes to discriminate whether an opinion is classified as positive or negative sentiment. The advantages of SVM implementation such as generalization, its stability, classification, and capability to process both linear and non-linear data have made SVM being considered as a reliable classification method for the research. According to these considerations, the authors carried out a sentiment analysis using SVM algorithm to identify the polarity scores of Go-Jek service reviews, which is mainly helpful for the stakeholders, particularly for Go-Jek.

The hybrid approach proposed on [33] to analyse the sentiment of Fintech users in Tegal City, especially the Ovo application which makes use of Support Vector Machine (SVM) and Particle Swarm Optimization (PSO) for classification of sentiments and Application RapidMiner Studio for data processing with an accuracy rate of 82 %.

[34] proposed an optimal Support Vector Machine for lung image classification where the parameters of SVM are feature selection and are managed by a modified grey wolf optimization algorithm combined with genetic algorithm (GWO-GA). A set of experiments has been performed for investigating the results in terms of feature selection results and classifier performance.

1.2.2) Neural Network (NN): A neural network is an algorithmic model which mimics the mode of how the human brain functions. In fact, neural networks denote to systems of neurons, either organic or artificial in nature and highly used in Sentiment Analysis. It simulates the neural structure of the brain having electronic networks of neurons. In this network, Neuron is the basic component and categorized in to three levels- input, hidden and output.

Sentiment Analysis of Indian Languages using Convolutional Neural Networks proposed on [35] Convolutional Neural Network for identifying the polarity scores of sentiments for better classification scores.

Aspect based sentiment analysis (ABSA) proposed on [36] uses a fine-grained sentiment analysis task, which comprises of two sub tasks two subtasks: aspect-category sentiment analysis (ACSA) and aspect-term sentiment analysis (ATSA). It also uses Long Short-Term Memory (LSTM) or Recurrent Neural Network (RNN), for training the data. A classic model based on gating mechanism, ensemble with convolutional neural networks (CNN) and self-attention mechanism was developed to test the sentiment scores of SemEval datasets for improving the accuracy rates.

To address this problem of effectively distinguishing the different polarities of variant targets in the same sentence, [37] proposed a deep neural network model which is an ensemble of convolutional neural network and regional long short-term memory (CNN-RLSTM) for the task of target-based sentiment analysis. The innovative approach reduces the training time of neural network model, controls the transmission of information through different weight matrices, and effectively deduce the sentiment polarities of different targets in the same sentence.

A novel neural network framework was proposed on [38] that combines recurrent and recursive neural models for aspect-based sentiment analysis for analysing the sentiment scores of User review datasets. Initially the datasets are divided into subsets using constituency and dependency parsers which include the sentiment information of aspect terms. Then the recursive neural trees are trained and the output is feed into the recurrent model. The proposed system achieved better state-of-the-art polarity results

Sentiment Analysis using Word2vec-CNN-BiLSTM Classification proposed on [39] uses a hybrid of Word Vector Model (Word2vec), Bidirectional Long-term and Short-term Memory networks (BiLSTM) and convolutional neural network (CNN) to achieve the accuracy rate of 91%.

1.2.3) Decision Tree classifier: Decision tree classifier provides a hierarchical decomposition of the training data

space subject to a condition on the attribute value that is used to divide the data. The condition or predicate is the occurrence or non-occurrence of one or more words. The division of the data space is performed recursively until the leaf nodes encompass certain minimum numbers of records which are castoff for the purpose of classification.

Sentiment Analysis of Restaurant Review with Classification Approach proposed in [40] uses Decision Tree-J48 Algorithm and WebHarvy for data Crawling in TripAdvisor dataset for restaurant recommendation system

Arabic Opinion Mining Using Parallel Decision Trees proposed on [41] uses Naïve Bayesian, and Support vector machine to classify opinion by applying parallel decision trees classifiers to have efficient results.

1.2.4)Rule-based classifiers:In rule-based classifiers, the data space is modelled with a set of rules. The left-hand side characterizes a condition on the feature set expressed in disjunctive normal form in Boolean logic, which consists of disjunction of conjunctions while the right-hand side characterises the class label. The conditions are on the term presence. Term absence is infrequently used since it is not informative in sparse data.

Fuzzy rule-based systems for interpretable sentiment analysis proposed on [42] performs classification of sentiments by using natural language processing, text analysis and computational linguistics for better polarity scores.

A Rule-based Expert System for Sentiment Analysis proposed on [43], focuses on text-based sentiment analysis using supervised classical for a unified framework for rule base and rule engine based out of Java The rules are designed based on standards and to enhance the knowledge base, cross-domain inputs and this system outperforms existing state-of-the-art classical approaches like Jess, IBM MS-BR and Drools

IV. SENTIMENT ANALYSIS OPEN PROBLEMS

The sentiment Analysis research area is highly prone to several open problems

A. Emoticons and Sarcastic Reviews

Emoticons are the pictorial representation of user's expressions. Using emoticons to express the product makes it easier for the customer or user to understand one's feelings. On the other hand, it becomes difficult for a machine to understand the emoticons. It is very tedious task to train an

algorithm for emoticons as an input. Sarcastic reviews which involve irony to mock or convey contempt, are difficult to interpret by the machine. The model wants to be trained with more and more such data to predict an accurate answer. Henceforth, Emoticons and sarcastic reviews are one of the biggest challenges and an active research area in sentiment analysis.

An Analysis on Sarcasm Detection Over Twitter During COVID-19 performed on [44] reveals that Sarcasm detection is one of the major challenges involved in Sentiment Analysis. Here major challenge is on extracting sarcastic content with negative content using Twitter Covid related tweets extracted on Twitter API during COVID-19 using linear support vector classifier (libSVM), Naïve Bayes, and Decision Tree for this analysis and yields 90% accuracy

B. Multilingual language Problem

As the dataset is a collection of random reviews contributed by the users, it can be in multiple languages. But the main challenge lies with the modelling of classifier which mainly trained to use with English language. Therefore, it becomes very cumbersome to train the algorithm for different or multiple languages other than English. Hence Multilingual Input is a big challenge in sentiment analysis.

C. False Data Problem

False or bogus reviews mislead the users or customers about a product by providing fake or irrelevant negative or positive reviews. This is generally performed to increase or decrease the popularity and acceptance of the product. So, identifying false reviews is a tedious and almost challenging task.

V. CONCLUSION

Through this paper, a comprehensive review on different techniques used for sentiment analysis were presented. The theoretical study includes the various trends, methodologies and practices to extract the sentiments from the textual data. From various classification methods, Sentiment Analysis classifies the textual data into positive, negative and neutral sentiments. The review paradigm explored various sentiment analysis algorithms that utilizes different Machine Learning classical algorithms like SVM, Naive Bayes, and neural networks as well as Lexicon based approaches and their working strategy. These classical models can be considered as the baseline learning methods which are very effective for Sentiment Analysis. This study also presented an overview on the recent advancements in Sentiment Analysis, real-life applications and open problems existed in the field. Using Natural Language Processing and Machine learning tools to reinforce the Sentiment Analysis process has attracted researchers recently and still an active research area.

REFERENCES

- [1] Tsytssarau Mikalai, Palpanas Themis. Survey on mining subjective data on the web. *Data Min Knowl Discov* 2012; 24:478–514
- [2] Yu Liang-Chih, Wu Jheng-Long, Chang Pei-Chann, Chu Hsuan-Shou. Using a contextual entropy model to expand emotion words and their intensity for the sentiment classification of stock market news. *Knowl-Based Syst* 2013;41:89–97.
- [3] Michael Hagenau, Michael Liebmann, Dirk Neumann. Automated news reading: stock price prediction based on financial news using context-capturing features. *Decis Supp Syst*; 2013.
- [4] Tao Xu, Peng Qinke, Cheng Yinzhaoh. Identifying the semantic orientation of terms using S-HAL for sentiment analysis. *Knowl-Based Syst* 2012;35:279–89.
- [5] Wilson T, Wiebe J, Hoffman P. Recognizing contextual polarity in phrase-level sentiment analysis. In: *Proceedings of HLT/ EMNLP*; 2005
- [6] Bilal Saberi, Saidah Saad. "Sentiment Analysis Or Opinion Mining: A Review", *International Journal of Advanced Science Engineering Information Technology*, Vol-7(2017), ISSN:2088-5334.
- [7] J. Bollen, H. Mao, and X. Zeng "Twitter mood predicts the stock market". *Journal of Computational Science*, 2(1): 1-8 2011
- [8] T. Xu, Q. Peng and Y. Cheng. "Identifying the semantic orientation of terms using S-HAL for sentiment analysis". *Knowledge-Based Systems*, 35: 279-289, 2012
- [9] Ahmed, Khaled, Neamat El Tazi, and Ahmad Hany Hossny. "Sentiment Analysis over Social Networks: An Overview." *Systems, Man, and Cybernetics (SMC)*, IEEE International Conference on. IEEE, 2015.
- [10] Arora, Piyush. " Sentiment Analysis for Hindi Language" *Diss.* International Institute of Information Technology Hyderabad, 2013.
- [11] Liu B. Sentiment analysis and opinion mining. *Synth Lect Human Lang Technol* 2012.
- [12] Diana Maynard, Adam Funk. Automatic detection of political opinions in tweets. In: *Proceedings of the 8th international conference on the semantic web, ESWC' 11*; 2011. p. 88–99
- [13] J. Bollen, H. Mao, and X. Zeng "Twitter mood predicts the stock market". *Journal of Computational Science*, 2(1): 1-8 2011
- [14] Miller G, Beckwith R, Fellbaum C, Gross D, Miller K. *WordNet: an on-line lexical database*. Oxford Univ. Press; 1990
- [15] Soumya S; Pramod K V, Fine Grained Sentiment Analysis of Malayalam Tweets Using Lexicon Based and Machine Learning Based Approaches, 4th Biennial International Conference on Nascent Technologies in Engineering (ICNTE), 2021
- [16] Rahmad Kumiawan;FitraLestari;AbdulSomadBatubara;MohdZakree Ahmad Nazri;KhairunnasRajab;Rinaldi Munir, Indonesian Lexicon-Based Sentiment Analysis of Online Religious Lectures Review, International Congress of Advanced Technology and Engineering (ICOTEN), 2021
- [17] Hu Mingming, Liu Bing. Mining and summarizing customer reviews. In: *Proceedings of ACM SIGKDD international conference on Knowledge Discovery and Data Mining (KDD'04)*; 2004.
- [18] Kim S, Hovy E. Determining the sentiment of opinions. In: *Proceedings of International conference on Computational Linguistics (COLING'04)*; 2004
- [19] Sandra Jardim;CarlosMora;Tiago Santana, A Multilingual Lexicon-based Approach for Sentiment Analysis in Social and Cultural Information System Data16th Iberian Conference on Information Systems and Technologies (CISTI),2021
- [20] SaprativaBhattacharjee;AnirbanDas;UjjwalBhattacharya;Swapan K. Parui;Sudipta Roy, Sentiment analysis using cosine similarity measure, IEEE 2nd International Conference on Recent Trends in Information Systems (ReTIS),2015
- [21] Yassin S. Mehanna;Massudi Bin Mahmuddin, A Semantic Conceptualization Using Tagged Bag-of-Concepts for Sentiment Analysis, Volume 9, IEEE,2021
- [22] Hatzivassiloglou V, McKeown K. Predicting the semantic orientation of adjectives. In: *Proceedings of annual meeting of the Association for Computational Linguistics (ACL'97)*; 1997.
- [23] Raihah Aminuddin; Aina Zuliana Zulkefli; Nor Aiza Moketar; Khyrina Airin Fariza Abu Samah, Sentiment analysis of online product reviews using Lexical Semantic Corpus-Based technique, IEEE, 2021
- [24] Yige Wang, Advanced Naive Bayes Algorithm Design with Part-of-Speech Tagger on Sentiment Analysis, International Conference on Computer Systems, Electronics and Control (ICCSEC),2017
- [25] Rosy Indah Permatasari; M. Ali Fauzi; Putra Pandu Adikara; Eka Dewi Lukmana Sari, Twitter Sentiment Analysis of Movie Reviews using Ensemble Features Based Naive Bayes, International Conference on Sustainable Information Engineering and Technology (SIET),2018
- [26] Siti Ernawati; Eka Rini Yulia; Friyadie; Samudi, Implementation of The Naive Bayes Algorithm with Feature Selection using Genetic Algorithm for Sentiment Review Analysis of Fashion Online Companies, 6th International Conference on Cyber and IT Service Management (CITSM), 2018
- [27] Kamal Sarkar ,Using Character N-gram Features and Multinomial Naive Bayes for Sentiment Polarity Detection in Bengali Tweets, Fifth International Conference on Emerging Applications of Information Technology (EAIT), 2018
- [28] Huihui Liang; Umarani Ganeshbabu; Thomas Thorne, A Dynamic Bayesian Network Approach for Analysing Topic-Sentiment Evolution, *Advanced Data Mining Methods for Social Computing*,2020
- [29] Yudi Chen;WenyinJi;Qi Wang ,A Bayesian-Based Approach for Public Sentiment Modeling, Winter Simulation Conference (WSC), 2019
- [30] Xiaodong Yan;TaoHuang,Tibetan Sentence Sentiment Analysis Based on the Maximum Entropy Model,10th International Conference on Broadband and Wireless Computing, Communication and Applications (BWCCA), 2015
- [31] Ahammed M S Shaik Afzal,Optimized Support Vector Machine Model for Visual Sentiment Analysis,3rd International Conference on Signal Processing and Communication (ICPSC), 2021
- [32] Windha Mega P.D.;Haryoko, Optimization Of Parameter Support Vector Machine (SVM) using Genetic Algorithm to Review Go-Jek's Services,4th International Conference on Information

- Technology, Information Systems and Electrical Engineering (ICITISEE), 2019
- [33] Warjiyono;SopianAji;Fandhilah;NunungHidayatun;HusniFaqih;Liesnaningsih, The Sentiment Analysis of Fintech Users Using Support Vector Machine and Particle Swarm Optimization Method,7th International Conference on Cyber and IT Service Management (CITSM),2019
- [34] I Jasmine Selvakumari Jeya and Dinesh Valluru, “IoT with Cloud Based Lung Cancer Diagnosis Model using Optimal Support Vector Machine”, Health care management science, Springer Publication, pp. 1-10, 2019.
- [35] Shalini K.;AravindRavikumar;Vineetha R C.;Aravinda Reddy D.;Anand Kumar M.Sentiment Analysis of Indian Languages using Convolutional Neural Networks, International Conference on Computer Communication and Informatics (ICCCI), 2018
- [36] Jian Yang;JuanYang,Aspect Based Sentiment Analysis with Self-Attention and Gated Convolutional Networks, IEEE 11th International Conference on Software Engineering and Service Science (ICSESS), 2020
- [37] Siyuan Chen;ChaoPeng;LinsenCai;Lanying Guo, A Deep Neural Network Model for Target-based Sentiment Analysis, International Joint Conference on Neural Networks (IJCNN), 2018
- [38] Cem RifkiAydin;Tunga Güngör, Combination of Recursive and Recurrent Neural Networks for Aspect-Based Sentiment Analysis Using Inter-Aspect Relations,IEEEAccess,Year: 2020
- [39] Jian Yang;Juan Yang, Aspect Based Sentiment Analysis with Self-Attention and Gated Convolutional Networks, IEEE 11th International Conference on Software Engineering and Service Science (ICSESS), 2020
- [40] Masrur Adnan;RiyanartoSarno;Kelly Rossa Sungkono, Sentiment Analysis of Restaurant Review with Classification Approach in the Decision Tree-J48 Algorithm, International Seminar on Application for Technology of Information and Communication (iSemantic), 2019
- [41] Wafa A. M. Ahmed;Alaa M. El-Halees,Arabic Opinion Mining Using Parallel Decision Trees, Palestinian International Conference on Information and Communication Technology (PICICT), 2017
- [42] Han Liu;MihaelaCocea, Fuzzy rule based systems for interpretable sentiment analysis,Ninth International Conference on Advanced Computational Intelligence (ICACI), 2017
- [43] Anuradha Yenikar;NarendraBabu;SunilSangve, R-SA: A Rule-based Expert System for Sentiment Analysis, IEEE Pune Section International Conference (PuneCon), 2019
- [44] Raju Kumar;Aruna BhatAn Analysis On Sarcasm Detection Over Twitter During COVID-19, 2nd International Conference for Emerging Technology (INCET), 2021