A Literature Survey On Sentiment Analysis Techniques Involving Social Media And Online Platforms

Raktim Kumar Dey, Debabrata Sarddar, Indranil Sarkar, Rajesh Bose, Sandip Roy

Abstract: Activities that take place or are influenced as a result of decisions being made are influenced by opinions at the root level. Analysis of opinions or sentiment analysis plays a vital role in trying to make as close approximation as possible. This is an extremely important aspect given that carefully planned and executed sentiment analyses can yield better and more accurate forecasts in politics as well as in business. At the base level, sentiment analysis stems from opinions shared or expressed by individuals and users. In the Internet space that permeates almost every known sphere and area of human activity on our planet, data in millions of bytes are posted and shared by individuals on social networking platforms, blogs, product review sites, and various other web forums. The potential to harvest such information and analyse the data can yield vital insights into how products, services, political personalities, companies, governments, etc. are perceived and viewed. Sentiment Analysis can engage multiple challenges such as accuracy-related issues, binary classification problem, data sparsity problem and polarity shift. While there have been various methods that have been postulated and developed for sentiment analysis, there yet remains to be an efficient approach in extracting and producing accurate sentiment analysis on a consistent basis. Although machine learning algorithms have come a long way, with Naïve Bayes, Support Vector Machine and Maximum Entropy being the significant ones to feature prominently in research and mainstream use, sentiment classification by category involving positive and negative sentiments, is a topic of research interest in its own right. This paper presents a survey on prominent Sentiment Analysis approaches and methodologies and seeks to submit a clear evaluation report upon which grounds for further research can be based.

Index Terms: Machine Learning, Naïve Bayes, Natural Language Processing (NLP), Neural Network, Opinion Mining (OM), Sentiment Analysis (SA), Social Media, Support Vector Machine (SVM).

---

1 INTRODUCTION

By nature, humans are said to be subjective and often biased with preconceived notions. Modern human societies are largely influenced by opinions in almost all spheres and domain of human civilization. Sentiment Analysis of opinions can, therefore, be viewed as an integral tool to analyse the mood and prevalent disposition of any sample group of individuals concerning any particular product, service, event or topic expressed in text form and published on social media platforms, blog posts, comments, web reviews, etc. With the proliferation of Internet and smart phones even in rural regions in underdeveloped countries around the world, opinions and reviews can be posted with ease. From the point of view of mining such data and opinionated text material, the challenge of examining all posts and reviews and assimilating such into meaningful orientation can be quite considerable. The idea of Sentiment Analysis (SA) has been, therefore, to table a summary of opinions segregated into positive, neutral and negative reviews based on analyses of texts posted by users in assorted digital platforms on the Internet [1] [2]. As an investigative approach to individual assessment and disposition of individuals towards events, topics, products, services, people and organizations, SA ranks as an important tool in this field [3].

- *Raktim Kumar Dey, Simplex Infra. Ltd., India, deyraktim@gmail.com*
- *Dr. Debabrata Sarddar, Univ. of Kalyani, India, dsarddar1@gmail.com*
- *Indranil Sarkar, Brainware University, India, indra.nil2004@gmail.com*
- *Dr. Rajesh Bose, Brainware University, India, bose.raj00028@gmail.com*
- *Dr. Sandip Roy, Brainware University, India, sandiproy86@gmail.com*

As a computational study into human attitudes, opinions and emotions, Sentiment Analysis (SA) or Opinion Mining (OM) ranks as being significant insofar as such human feelings are expressed in terms of any entity. An entity can be said to represent events, individuals or even topics. To extract sentiment, SA uses three terms. These are object and feature, opinion holder, and opinion and orientation. Natural Language Processing (NLP), text analyses and computational methods to automatically extract and classify sentiments from reviews posted online form the chief pillars on which SA is based [4]. The task of undertaking SA is not easy with several technical hurdles that can be thrown in the way. These hurdles can range from identifying objects, opinion orientation classification through to feature extraction. SA is usually carried out either by supervised learning or unsupervised learning techniques. Unsupervised learning techniques involve Naïve Bayes, Neural Networks and Support Vector Machine design approaches. Among these, SVM is considered most appropriate for carrying out SA. Assumption investigation, that is considered a part of SA, is needed to be carried out at three levels involving the aspect level, document level and sentence level [4]. SA is also known as opinion extraction, opinion mining, effect analysis, sentiment mining, emotion analysis, review mining, etc. [5]. Although the terms are synonymous, depending on the nature of tasks, their intrinsic meanings and definitions may vary ever so slightly [5]. Although, SA and OM are terms that are often deemed as interchangeable given their inherently common definition root and applicability in the real world of studying human attitudes and opinions, some researchers are of the view that these are not quite exactly the same [6]. Such group of researchers have opined that while OM extracts and then helps in analysing human opinions on any topic, person, organization, product, service, etc., SA goes beyond merely identifying sentiments expressed in text form by analysing text extracted. Thus, such select group of researchers postulate that the purpose of SA is to detect opinions, isolate and identify underlying sentiments and then go on to classify polarity of the opinions examined. SA has found global acceptance as an important tool in analysing data to help predict election results [7], forecast stock market positions of businesses [8] and help estimate sales of products across diverse market conditions [9]. In conducting SA, it is often assumed that the data analysed would have opinions expressed explicitly. In most cases, however, only facts and, as such, objective information only is presented in documents that are being analysed. A case in point would be news items. In other cases, documents that are assumed to contain opinions or sentiments may also include sentences that present facts. It becomes clear as such, that the most significant feature of SA would be to first identify nature and type of sentences in text that is being analysed. Classifying sentences in text as either objective or subjective is a key task that needs to be performed while conducting a SA.

2 CLASSIFICATION LEVELS

Sentiment Analysis (SA) is considered a three-layered approach. The first of the layers is document based. The second being sentence based. The third is the aspect level, also considered as word or phrase based.

2.1 Document Level

In SA, the first level is considered the document level. At this level, an entire document is considered as a whole for SA. In this treatment for conducting SA, the one making an opinion is considered a single source or an individual entity [10]. A different aspect of SA at this level has been observed to be sentiment regression [11], [12]-[15]. To gauge the extent of positive or negative outlook, many researchers have turned to using supervised learning to estimate ratings of a document [11]. In another study, researchers have proposed a linear-based combination approach based on polarities observed in text documents [10][16]. A major problem observed while dealing at document level is at this layer not all sentences involving expression of opinions can be deemed as subjective sentences. Hence, accuracy of results is dependent on how closely each sentence is extracted and analysed individually. This method, therefore, promotes the rate at which subjective sentences can be extracted for the purpose of SA over objective sentences that can be set aside. Research in SA techniques often place major thrust and emphasis at the sentence level.

2.2 Sentence Level

Research studies abound on classifying and analysing each sentence in a document or piece of text as either objective or subjective ones. In one example, the authors [17] propose conducting SA on subjective sentences alone following classification. As a tool for detecting subjective sentences, machine learning has been an asset for researchers and scholars in this field. In one study [18], the authors have proposed a model based on logarithmic probability rate and a number of root terms to form the basis of scorecard for categorization of each subjective sentence classified. In another paper [19], research scholars have postulated a model that takes into account sentiments of all terms in each sentence to formulate an overall sentiment for a sentence that is under consideration. SA at sentiment level is not without its own shortcomings. There may be objective sentences that may actually have sentiments not detected. An example of such a sentence could be: “I bought a table from a reputed online store only to find that its legs are not stable enough.” This is a typical example of what a review could be like posted on any of the prominent online retail websites. While the sentence presents certain facts or observations, it also infuses a sentiment indirectly. The reviewer or poster of the comment has aired a negative outlook. To offset such kind of SA challenges, analyses at word or phrase level need to be conducted.

2.3 Aspect Level

Aspect Level SA aims at addressing the shortcomings of document and sentence levels of SA. Fine-grained control can be exercise with the help of SA. The target focus of aspect level SA is to examine opinion in critically and exclusively. Aspect level SA assumes that an opinion can only be one among positive, neutral, negative or an outright objective sentiment expressed [20][21]. If one were to consider the sentence “Telephone call quality of Sony phones are remarkable save and except for the quality of battery”, two statements are immediately apparent. The first being that the call quality of Sony phones is good, while that of battery is not so much. This approach, therefore, enables turning unstructured content into organized form of information that can be subjected to a range of subjective and quantitative experiments. Such finer degree of SA is mostly beyond the scope of document and sentence level analyses.

3 SENTIMENT CLASSIFICATION TECHNIQUES

Sentiment classification techniques can be segregated into three categories (Fig. 1.). These are machine learning, lexicon-based and hybrid approaches [22]. The first of these techniques involve popular machine learning (ML) algorithms and involves using linguistic features. The second involves analyses through a collection of sentiment terms that are precompiled into sentiment lexicon. This is further divided into dictionary- and corpus-based approaches that use semantic or statistical methods to gauge the extent of polarity of sentiment. The hybrid approach involves combining ML and lexicon-based approaches. The following illustration aims at providing an insight into more popular algorithms used in sentiment classification techniques.

3.1 Machine Learning Approach

In machine learning approach, machine learning (ML) algorithms are used almost exclusively and extensively to conduct SA. ML algorithms are used on conjunction with linguistic and syntactic features [23, 24, 25, 26].

3.1.1 Supervised Learning

Supervised Learning involves datasets that are clearly labelled. Such datasets are share with assorted supervised learning models [23, 24, 25, 27].

3.1.1.1 Decision Tree Classifiers

This type of classifier extends a hierarchical breakdown of training data space in which attribute values are used for data segregation [28]. This method is predicated or based upon the absence or presence of one or more words and is conducted recursively until a minimum number of records are registered with lead nodes that are used for classification.

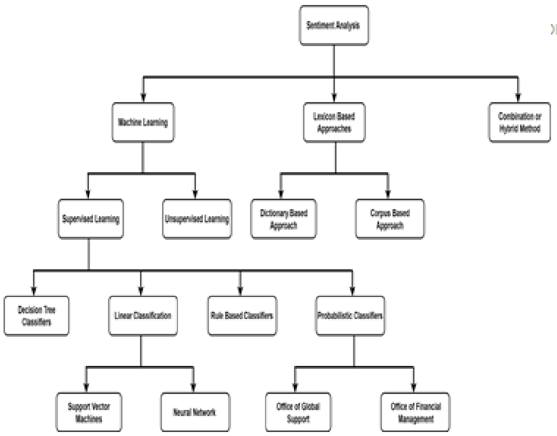

Fig 1*. Sentiment classification techniques.*

3.1.1.2 Linear Classification
Linear classification models involve support vector machine (SVM). This is a form of classifier that is focused on segregating and isolating direct separators between different classes. It also involves neural system [29]. SVM is a form of supervised learning model and works chiefly on the principle of decision boundaries setup by decision planes. A decision plane is defined as a set of objects that are part of a range of class memberships.

3.1.1.2.1 Support Vector Machines Classifiers (SVM)
SVMs have been designed to fundamentally identify and isolate linear separators in search space for the purpose of categorizing assorted classes. Ideally, text data is considered suitable for SVM classification in view of sparse nature accorded by text. Since, only a select number of features are irrelevant notwithstanding the tendency to be correlated and organized into linear distinguishing buckets, text data is often considered an ideal candidate for SVM [33]. With the help of SVM, a non-linear decision surface can be constructed out of original feature space. This can be achieved through mapping of data instances in a non-linear fashion to an inner product space where classes can be linearly segregated using hyperplane [34].

3.1.1.2.2 Neural Network (NN)
Neural Networks are constituted of, as the term suggests, by neurons as the basic fundamental building block. Inputs to neurons are depicted using a vector and denoting word frequencies in a document across a line. A set of weights are considered for each neuron to enable computation of a function of the inputs involved. For boundaries that are non-linear, multilayer neural networks are employed. These multiple layers are used in conjunction with multiple pieces of linear boundaries used for approximation of enclosed regions involving particular class. Neuron outputs generated in previous layers are used to feed the neurons in subsequent layers. The training process, therefore, becomes progressively complex as errors are reversing propagated across all layers [35, 36]. As shown by the authors [37], SVM and NN can also be deployed for classifying relationships that are of personal nature in biographical texts. In their research, relations were marked as being positive, neutral or unknown between two individuals. The case study involved historical information on biographies of individual of a given region and year. Their research revealed that classifiers showed promising results in identifying relations. The research also revealed that training set that contained relationships involving surrounding individuals, tended to produce more accurate results than a training set that involved only a specific individual or an entity. In terms of scoring accuracy and as a benchmark of practicability, the authors were able to demonstrate that SVM and a single layer NN (1-NN) algorithm together were effective in attaining highest scores.

3.1.1.3 Rule-based Classifiers
Data space is modelled with a set of rules in rule-based classifiers. On the left-hand side, feature set is expressed in disjunctive normal form to denote condition of the feature set itself. While on the right hand side, class label is inserted. Conditions are usually formulated on basis of presence of term. The opposite, or term absence, is not normally invoked as it does not project informative behaviour on sparse data. Several criteria are used to generate rules and training phases create rules based on criteria chosen. The two most common are being confidence and support [30]. Support criterion is considered to be the number of instances in training data set relevant for the rule. Confidence criterion involves conditional probability and denotes that the right hand side of the rule is satisfactorily met only upon satisfying the left hand side. The authors [31] have proposed combined rule algorithms. While decision rules and decision trees often involve encoding rules in feature space, the latter attempts to reach the goal using a hierarchical approach. In a research [32], it was observed that a certain path in decision tree can be identified as being appropriate for classification rule in a text instance. The primary difference between the two is that while decision trees involve hierarchical partitioning of data space that is considered rigid, the latter is more flexible and can accommodate overlapping in decision space.

3.1.1.4 Probabilistic Classifiers
This type of classifier involves mixing of models to achieve classification. This model assumes that every class involved is a component of the mixture itself. Each component is a generative model that functions as probability sampler of any given term for that component. Alternatively, such kinds of classifiers are also known as generative classifiers. Three important probabilistic classifiers stand out. These are Bayesian Network, Naïve Byes, and Maximum Entropy [3, 5, 39].

3.1.1.4.1 Naïve Bayes Classifier (NB)
This form of classification is used almost extensively for classifying text documents and conducting SA on such form of documents [40][41][42][43]. The technique is founded on a probabilistic approach and uses cooperative probabilities of specific terms with a text document as an input for approximation of probability of a certain group.

3.1.1.4.2 Bayesian Network
The Naïve Bayes classifier assumes that each and every component is complete independent of the other. This enables BN to display an acyclic graph that is not only coordinated but also relates random variables with edges representing dependencies that are in turn conditional. BN approaches and examines factors and associations that exist between them. In

such manner, an entire probability joint distributed over each element can be resolved. Considering that computational complexity of BN is expensive in terms of text mining, it is sporadically used.

3.1.1.4.3 Maximum Entropy Classifier (ME)
A type of probabilistic classifier with a place among exponential class of models, it does not rely on the assumption that components are independent. On the other hand, ME is dependent on Principle of Maximum Entropy. It selects the model that has the largest entropy. ME classifiers find use in applications involving dialect identification, assumption investigation, point arrangement, etc.

3.1.2 Unsupervised Techniques
In this form of approach, classification of sentiment is achieved through comparison. Component comparisons take place involving word lexicons that are assigned sentiment values before use [3, 44]. The more popular forms of such group of techniques are hierarchical and partial clustering [45, 46].

3.2 Lexicon-based Approach
This kind of approach involves determination of polarity by employing opinion words from sentiment dictionary and matches those with data. Such an approach marks sentiment scores to indicate positive, negative or objective types of words. Lexicon-based approaches are dependent on sentiment of lexicon involving a set of precompiled and known sentiment phrases, terms and idioms. Two forms of sub-classifications exist for this type of approach. These are discussed in subsequent sections.

3.2.1 Dictionary-based Approach
In this approach, arrangement of words is made possible through manual approach involving a group of instructions that are known beforehand. The conclusion set is generated by looking up a notable corpus – WordNet, for appropriate words and antonyms relevant for the SA [14, 15]. The process is iterative and stops only when no new words are detected. Else, subsequent iterations follow when words are progressively appended to seed list. After the process stops, a manual appraisal is conducted to evaluate and correct errors. However, this approach is not without its flaws as it is often unable to detect in certain circumstances involving specific introductions or words with spaces [48, 49].

3.2.2 Corpus-Based
This approach involves dictionaries specific to given domain. The dictionaries are produced on the basis of seeds of opinion terms that grow out of search for related words through use of statistical or semantic procedures.

3.3 Combination or Hybrid Method
Aside from involving individual ML approaches or Lexicon-based approach that has been described earlier, there are select few research techniques that involve a mixture of both. The improved Naïve Bayes and SVM algorithms find frequent mention in research studies [50]. To narrow the gap between positive and negative, feature selections such as unigrams and bigrams are often used. Studies have shown that combination of ML and dictionary-based methods can significantly improve the level of classification of sentiments.

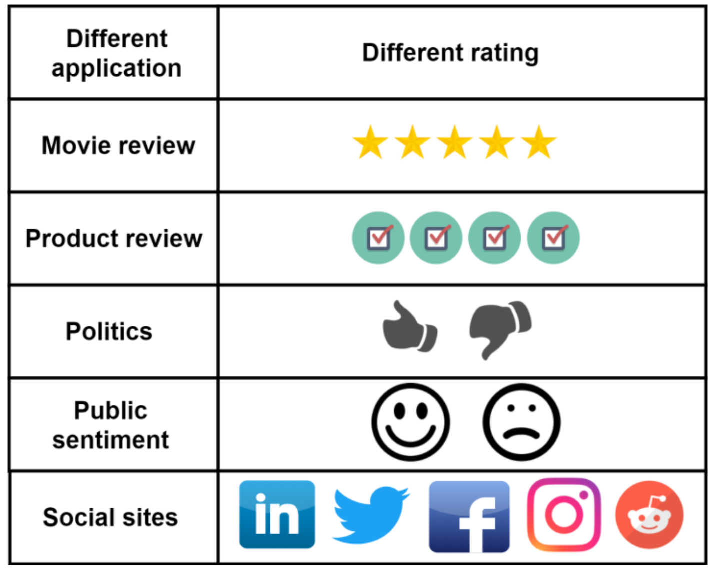

Fig 2. *Application of sentiment analysis.*

4 APPLICATION OF SENTIMENT ANALYSIS
Applications for SA are wide-ranging. SA can be used in cases involving politics, product reviews, cinema reviews, social media posts and much more [5, 20, 39]. The fig. 2 shows the form of application of SA in the context of reviews of cinemas [20]. In a broader and more global context, SA has been thrust into limelight because of the potential to gauge more accurately and extensively data that can be harvested from social media platforms, e.g., Twitter, Facebook, LinkedIn, etc. [39].

5 LITERATURE REVIEW
Kouloumpis et al. [51] highlighted the efficacy of existing lexical resources and linguistic features for conduct SA on Twitter messages and similar micro-blogging posts. The researchers postulated that micro-blogging is more relevant and appropriate in comparison to part-of-speech features and those belonging to existing sentiment lexicon. The authors concluded that inclusion of microblogging features is not likely to augment training data. Hybrid classification was also found to be more relevant and showed promising results in another research [52] involving rule-based classification and supervised learning processes. In the context of natural language processing (NLP), SA has grown in stature as one of the most researched topics since the turn of the century [53, 54]. Researchers have been constantly examining sentences and assorted types to streamline SA methods. Concept level SA system (psenti) as has been postulated by Mudinas et. al. have been shown to be more effective than lexicon-based system [55]. Results obtained from their research reveal that hybrid approach is considerably more efficient and returned more accurate results. In their research [12], Tripathy et. al. proposed four ML algorithms for sentiment classification. These were Naïve Bayes, Maximum Entropy, Stochastic Gradient Descent and Support Vector Machine. Their research demonstrated that accuracy can be achieved through progressive classification. Their research was conducted on popular movie review website – IMDB. Deploying an n-gram approach, the authors were able to produce consistent high accuracy using a combination of TF-IDF and count vectorizer

technique. To counter the problem of words and punctuation symbols that although were understandable to humans, had no formal definition in the English lexicon, the authors developed a new list of such words to assist in SA. Mitchel et al. modeled sentiment detection in their research paper. The focus of their work was to show applicability of sentiment detection as a problem that involved sequence tagging [56]. Their work was expanded upon by later research work [57] that examined embedding of words and automated combination of features through neural networks. Arun et. al. proceeded to conduct SA on tweets involving demonetization in Indian economy. Their research approach [58] was to extract data from Twitter and convert such into text. This text was to act as input dataset. SA was then performed following removal of stop words so that determination of polarity of the words could be carried out and the actual tweets, therefore, could be identified as either positive or negative. A new method was postulated for SA on demonetization and its subsequent effects and the wave of public opinion that it had unleashed. Bigrams, data cleaning, polarity, sentiment scores and graphical methods were all used in the research. In conducting a comparative analysis of assorted approaches for SA and topic detection, researchers [59] examined a collection of tweets in Spanish. Lemmatizers, stemmers, n-grams, negations, valence shifters, Twitter hashtags and semantics were explored and presented in a detailed study. The authors opined that lack of context and extreme short nature of text involved, tweets are difficult to make clear assessment of. In another research study involving e-commerce and online product reviews, the authors [60] stated that public opinions and those of buyers of products sold on online retail platforms would have far-reaching significance in terms of profitability and economic viability. As such, entrepreneurs and enterprises invest in opinion mining and SA to work out ways to stay ahead of the competition. The authors use fuzzy rule-based systems (FRBS) with Mamdani and Takagi Sugeno Kang (TSK) models deployed in FRBAS R package. The authors go on to compare these models with alternate classification techniques and grade them on the lines of precision, recall, f-measure, performance and accuracy. Exploring the feasibility and efficacy of SA in the context of Arabic tweets, the authors [61] have observed that there is a significant lack of resources to carry out SA in Arabic. The authors conducted their research using Jordanian Arabic tweets and examined such using a range of supervised ML approaches. Their research revealed that SVM classifier with TF-IDF and bigrams was comparatively better than Naïve Bayesian one. For their experiment, the authors had collected close to a couple of thousand Arabic tweets posted in Jordanian dialect. The two ML algorithms of SVM and NB were then run using a range of ngrams with varying weighting schemes and following application of stemming techniques. In conclusion, the authors observed that SVM classifier using stemmer and TF-IDF weighing than any other known Arabic SA research observations. A novel approach was adopted by Socher et. al. when the authors [62] proposed a semi-supervised recursive auto-encoding approach for forecasting sentiment distributions without the benefit of sentiment lexica or rules that involved shifting of polarity. In their paper [63], the authors proposed recursive neural tensor network (RNTN) to analyse and discern phrases and sentences of varying length. In another paper [64], an alternative form of sentiment classification was proposed to analyse sentences of different lengths and words. While hyper parameter tuning for word vectors has been the research of choice [65], deep recursive neural network (DRNN) has been selected by the authors [66] for conducting sentiment classification assignments. Interestingly, the authors [1] postulate that for obtaining improved classification results, larger datasets involving crowd sourcing and semi-supervised learning held far greater potential. The authors used Naïve Bayes, Support Vector Machine and K-Nearest Neighbour to conduct comparison tests. Their experiments revealed that while SVM was most useful in turning in highest precision, it was KNN that achieved highest recall consistently.

6 CONCLUSION

Sentiment Analysis is now by and large considered to be critical in terms of socio-economic standpoint. Understanding SA and examining the methods that can help achieve accuracy in a wide variety of input formats is critical for businesses, institutions and individuals to survive and succeed. This survey paper presents the applications and challenges that lie ahead of sentiment classification in single and cross domains. Before SA can be conducted, subjectivity analysis, negation handling and polarity classification are required to be conducted. This paper also discusses selected supervised and unsupervised methods that are of significance.

REFERENCES

[1]. S. ChandraKala, and C. Sindhu, "Opinion mining and sentiment classification: A survey", ICTACT journal on soft computing, vol. 3, No. 1, pp. 420-425, October 2012.

[2]. B. Pang, and L. Lee, "Opinion mining and sentiment analysis", Foundations and Trends® in Information Retrieval, Vol. 2, No. (1–2), pp. 1-135, July 2008, doi:10.1561/1500000011.

[3]. E. Aydoğan, and M.A. Akcayol, "A comprehensive survey for sentiment analysis tasks using machine learning techniques", In Proceedings of 2016 International Symposium on INnovations in Intelligent SysTems and Applications (INISTA) , Sinaia, Romania, pp. 1-7, August 2016.

[4]. B. Agarwal, N. Mittal, P. Bansal, and S. Garg, "Sentiment analysis using common-sense and context information", Computational intelligence and neuroscience, Vol. 2015, March 2015, doi:10.1155/2015/715730.

[5]. W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey", Ain Shams engineering journal, Vol. 5, No. 4, pp. 1093-1113, December 2014, doi:10.1016/j.asej.2014.04.011.

[6]. M. Tsytsarau, and T. Palpanas, "Survey on mining subjective data on the web", Data Mining and Knowledge Discovery, Vol. 24, No. 3, pp. 478-514, May 2012, doi:10.1007/s10618-011-0238-6

[7]. J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market", Journal of computational science, Vol. 2, pp. 1-8, March 2011, doi:10.1016/j.jocs.2010.12.007.

[8]. T. Xu, Q. Peng, and Y. Cheng, "Identifying the semantic orientation of terms using S-HAL for sentiment analysis", Knowledge-Based Systems, Vol. 35, pp. 279-289, September 2012,

doi: 10.1016/j.knosys.2012.04.011.

[9]. J. Brooke, M. Tofiloski, and M. Taboada, "Cross-linguistic sentiment analysis: From English to Spanish.", In Proceedings of International Conference RANLP-2009 , Borovets, pp. 50-54, September 2009,

[10]. J. Kamps, M. Marx, R.J. Mokken, and M. De Rijke, "Using WordNet to measure semantic orientations of adjectives", In LREC , Lisbon, Portugal , Vol. 4, pp. 1115-1118, May 2004.

[11]. B. Liu, M. Hu, and J. Cheng, "Opinion observer: analyzing and comparing opinions on the web", In Proceedings of the 14th international conference on World Wide Web, pp. 342-351, May 2005, doi:10.1145/1060745.1060797.

[12]. L. Qu, G. Ifrim, and G. Weikum, , " The bag-of-opinions method for review rating prediction from sparse text patterns", In Proceedings of the 23rd international conference on computational linguistics, Coling 2010, Beijing ,Association for Computational Linguistics, pp. 913-921, August 2010.

[13]. M.D. Molina-González, E. Martínez-Cámara, M.T. Martín-Valdivia, and J.M. Perea-Ortega, "Semantic orientation for polarity classification in Spanish reviews", Expert Systems with Applications, Vol. 40, No. 18, pp. 7250-7257, 2013, doi:10.1016/j.eswa.2013.06.076.

[14]. P. Chaovalit, and L. Zhou, "Movie review mining: A comparison between supervised and unsupervised classification approaches", In Proceedings of the 38th annual Hawaii international conference on system sciences, pp. 112c-112c, January 2005, doi: 10.1109/HICSS.2005.445.

[15]. Y. Chen, and J. Xie, "Online consumer review: Word-of-mouth as a new element of marketing communication mix", Management science, Vol. 54, No. 3, pp. 477-491, March 2008, doi: 10.1287/mnsc.1070.0810.

[16]. Balahur, R. Mihalcea, and A. Montoyo, "Computational approaches to subjectivity and sentiment analysis: Present and envisaged methods and application", Computer Speech & Language, Vol. 28, No. 1, pp. 1-6, January 2014, doi: 10.1016/j.csl.2013.09.003.

[17]. Aue, and M. Gamon, "Customizing sentiment classifiers to new domains: A case study", In Proceedings of recent advances in natural language processing (RANLP), pp. 2.1, 2005.

[18]. J. Brooke, M. Tofiloski, and M. Taboada, "Cross-linguistic sentiment analysis: From English to Spanish", In Proceedings of the international conference RANLP-2009, Borovets, pp. 50-54, September 2009.

[19]. R. Bose, R.K. Dey, S. Roy, and D. Sarddar, "Analyzing Political Sentiment Using Twitter Data", In Information and Communication Technology for Intelligent Systems , Springer, Singapore, pp. 427-436, 2019.

[20]. X. Fang, and J. Zhan, "Sentiment analysis using product review data", Journal of Big Data, Vol. 2, No. 1, p.5, June 2015, doi: 10.1186/s40537-015-0015-2.

[21]. V. A. Kharde, and S. Sonawane, "Sentiment Analysis of Twitter Data: A Survey of Techniques" , International Journal of Computer Applications, Vol. 139, No. 11, pp. 5-15, April 2016, doi:10.5120/ijca2016908625.

[22]. D. Maynard, and A. Funk, "Automatic detection of political opinions in tweets", In Proceedings of the 8th international conference on the semantic web, ESWC[1] 11, pp. 88–99, May 2011.

[23]. W. Meddhat , Ahmed Hassan ,Hoda Korashy "Sentiment analysis algorithms and applications: A surve", Ain Sham University, Faculty of Engineering, Computer & Systems Department, Egypt 19, Vol. 5, No. 4, pp. 1093-1113, December 2014.

[24]. E. Erdogan, and M. A. Akyol, "A Comprehensive Survey for Sentiment Analysis Tasks Using Machine Learning Techniques", In proc. 2016 International Symposium on INnovations in Intelligent SysTems and Applications (INISTA), IEEE, Sinaia, Romania, August 2016, doi: 10.1109/INISTA.2016.7571856.

[25]. F. Neri, C. Aliprandi, F. Capeci, M. Cuadros, T. By, "Sentiment Analysis on Social Media", In Proceedings of 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, Istanbul, August 2012,doi: 10.1109/ASONAM.2012.164.

[26]. P. Chikersal, S. Poria, and E. Cambria "SeNTU: Sentiment Analysis of Tweets by Combining a Rule-based Classifier with Supervised Learning", In Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), Denver, Colorado, June 2015, pp. 647–651.

[27]. L. I. Tan, W.S. Phang, K.O. Chain, "Rule-based Sentiment Analysis for Financial News", In Proc. Systems, Man, and Cybernetics (SMC), 2015 IEEE International Conference, Kowloon, China, January 2016, doi: 10.1109/SMC35812.2015.

[28]. J.R. Quinlan, "Induction of decision trees", Machine learning 1, pp.81-106, March 1986, doi: 10.1007/BF00116251.

[29]. Sharma, and Aakanksha, "A Comparative Study Of Sentiments Analysis Using Rule Based and Support Vector Machine", International journal of Advanced Research in Computer and Communication Engineering, Vol. 3, No. 3, March 2014.

[30]. Liu, W. Hsu, and Y. Ma, "Integrating classification and association rule mining", In Proc. ACM Int. Conf. on Knowledge Discovery and Data Mining (SIGKDD '98), New York City, Vol. 98, pp. 80-86, August 1998.

[31]. W. Medhat, A. Hassan, and H. Korashy, "Combined algorithm for data mining using association rules", Ain Shams Journal of Electrical Engineering, Vol. 1, No. 1, October 2014.

[32]. K. Chandra, P. Bhattacharjee, S. Roy, and S. Biswas, "Intelligent Data Prognosis of Recurrent of Depression in Medical Diagnosis", In Proc. International Conference on Reliability, Infocom Technologies and Optimization (ICRITO 2020), India, June 2020.

[33]. T. Joachims, "A Probabilistic Analysis of the Rocchio Algorithm with TFIDF for Text Categorization", No. CMU-CS-96-118, Carnegie-mellon univ, Pittsburgh, pa: dept. of computer science, 1996.

[34]. M.A. Aiserman, E.M. Braverman, and L.I. Rozonoer, "Theoretical foundations of the potential function method in pattern recognition", Avtomat. i

Telemeh, Vol. 25, No. 6, pp. 917-936, 1964.
[35]. M.E. Ruiz, and P. Srinivasan, “Hierarchical neural networks for text categorization”, In Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval , pp. 281-282, August 1999.
[36]. H.T. Ng, W.B. Goh, and K.L. Low, “Feature selection, perceptron learning, and a usability case study for text categorization”, In Proceedings of the 20th annual international ACM SIGIR conference on Research and development in information retrieval(SIGIR’ 97), Philadelphia USA, pp. 67-73, July 1997.
[37]. M.V.D. Camp, A.V.D. Bosch, “The socialist network”, Decision Support Systems, ELSEVIER, Vol. 53, No. 4, pp. 761-769 · November 2012.
[38]. H. Kang, S.J. Yoo, and D. Han, ”Senti-lexicon and improved Naïve Bayes algorithms for sentiment analysis of restaurant reviews”, Expert Systems with Applications, Vol. 39, No. 5, pp. 6000-6010, 2012.
[39]. Batrinca, and P.C. Treleaven, “Social media analytics: a survey of techniques, tools and platforms”, Ai & Society, Vol. 30, No. 1, pp. 89-116, 2015.
[40]. 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, 2011.
[41]. P. Chaovalit, and L. Zhou, ”Movie review mining: A comparison between supervised and unsupervised classification approaches”, In Proceedings of the 38th annual Hawaii international conference on system sciences, Big Island, HI, USA, pp. 112c-112c, January 2005.
[42]. Z. Zhang, Q. Ye, Z. Zhang, and Y. Li, "Sentiment classification of Internet restaurant reviews written in Cantonese", Expert Systems with Applications, Vol. 38, No. 6, pp. 7674-7682, 2011.
[43]. Agarwal, B. Xie, I. Vovsha, O. Rambow, and R.J. Passonneau, “Sentiment analysis of twitter data”, In Proceedings of the Workshop on Language in Social Media (LSM 2011), Portland, Oregon, pp. 30-38, June 2011.
[44]. W. Medhat, A. Hassan, and H. Korashy, ”Sentiment analysis algorithms and applications: A survey”, Ain Shams engineering journal, Vol. 5, No. 4, pp. 1093-1113, 2014.
[45]. F. Neri, C. Aliprandi, F. Capeci, M. Cuadros, and T. By, “Sentiment analysis on social media”, In Proc. 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, Istanbul, Turkey, pp. 919-926, August 2012.
[46]. Sarddar, R. K. Dey, R. Bose, and S. Roy, “Topic Modeling as Tool to Gauge Political Sentiments from Twitter Feeds”, International Journal of Natural Computing Research, Vol. 9, No. 2, pp. 1-22, 2020.
[47]. R. Bose, R. K. Dey, S. Chakraborty, S. Roy, and D. Sarddar, “Examining Hidden Meaning of E-commerce Platform”, International Journal of Innovative Technology and Exploring Engineering, Vol. 8, No. 12, pp. 257-261, 2019.
[48]. C.N. Chandra, S. Gupta, and R. Pahade, “Sentiment analysis and its challenges”, International Journal of Engineering Research & Technology, Vol. 4, No. 03, pp. 968-970, 2015.
[49]. V. Kharde, and P. Sonawane, “Sentiment analysis of twitter data: a survey of techniques”, International Journal of Computer Applications, Vol. 139, No. 11, pp. 5-15, April 2016.
[50]. P. Arora, A. Bakliwal, and V. Varma, “Hindi subjective lexicon generation using WordNet graph traversal”, International Journal of Computational Linguistics and Applications, Vol. 3, No. 1, pp. 25-39, June 2012.
[51]. Kouloumpis, T. Wilson, and J. Moore, “Twitter sentiment analysis: The good the bad and the omg!”, In Proceedings of the Fifth International AAAI conference on weblogs and social media, Barcelona, Catalonia, Spain, pp. 538-541, July 2011.
[52]. R. Prabowo, and M. Thelwall, “Sentiment analysis: A combined approach”, Journal of Informetrics, Vol. 3, No. 2, pp. 143-157, 2009.
[53]. B. Liu, “Sentiment analysis: Mining opinions, sentiments, and emotions”, Cambridge University Press, June 2015.
[54]. K. Ravi, and V. Ravi, ”A survey on opinion mining and sentiment analysis: tasks, approaches and applications”, Knowledge-Based Systems, ELSEVIER, Vol. 89, pp. 14-4, November 2015.
[55]. Mudinas, D. Zhang, and M. Levene, “Combining lexicon and learning based approaches for concept-level sentiment analysis”, In Proceedings of the first international workshop on issues of sentiment discovery and opinion mining, Beijing China, pp. 1-8, August 2012, doi:10.1145/2346676.2346681.
[56]. M. Mitchell, J. Aguilar, T. Wilson, and B. Van Durme, “Open domain targeted sentiment”, In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, Seattle, Washington, USA, pp. 1643-1654, October 2013.
[57]. M. Zhang, Y. Zhang, and D.T. Vo, “Neural networks for open domain targeted sentiment”, In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pp. 612-621, September 2015.
[58]. K. Arun, A. Srinagesh, and M. Ramesh, ”Twitter sentiment analysis on demonetization tweets in India using R language”, International Journal of Computer Engineering In Research Trends, Vol. 4, No. 6, pp. 252-258, June 2017.
[59]. A.F. Anta, L.N. Chiroque, P. Morere, and A. Santos,”Sentiment analysis and topic detection of Spanish tweets: A comparative study of of NLP techniques”, Procesamiento del lenguaje natural, 50, pp.45-52, April 2013.
[60]. S.M. Basha, Y. Zhenning, D.S. Rajput, N. Iyengar, and D.R. Caytiles, “Weighted fuzzy rule based sentiment prediction analysis on tweets”, International Journal of Grid and Distributed Computing, Vol. 10, No. 6, pp. 41-54, June 2017.
[61]. K.M. Alomari, H.M. ElSherif, and K. Shaalan, “Arabic tweets sentimental analysis using machine learning”, In International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems, pp. 602-610, Springer, June 2017, doi:10.1007/978-3-319-60042-0_66.
[62]. R. Socher, J. Pennington, E. H. Huang, A. Y. Ng, & C. D. Manning, “Semi-supervised recursive autoencoders for predicting sentiment distributions”, In Proceedings of the conference on empirical

methods in natural language processing, pp. 151–161, , July 2011.

[63]. R. Socher, A. Perelygin, J. Wu, J. Chuang, C.D. Manning, A.Y. Ng, and C. Potts, "Recursive deep models for semantic compositionality over a sentiment treebank", In Proceedings of the 2013 conference on empirical methods in natural language processing, Seattle, Washington, USA, pp. 1631-1642, October 2013.

[64]. N. Kalchbrenner, E. Grefenstette, and P. Blunsom, 2014. "A convolutional neural network for modeling sentences", In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Baltimore, Maryland, pp. 655–665, June 2014, doi: /10.3115/v1/P14-1062.

[65]. Y. Kim, "Convolutional neural networks for sentence classification.", In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP) , Doha, Qatar, August 2014, doi:/10.3115/v1/D14-1181.

[66]. O. Irsoy, and C. Cardie, "Opinion mining with deep recurrent neural networks", In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), pp. 720–728, October 2014.