Sentiment Analysis Methods using Lexicon Approach

Surendra Kumar[1]*, Suryakant Pathak[2]

[1]Research Scholar, Department of Computer Science and Engineering, Dr.K.N.Modi University, Newai, Rajasthan, India
[2]Professor, Department of Computer Science and Engineering, Dr.K.N.Modi University, Newai, Rajasthan, India

Abstract

Opinions plays an important role in our daily life. We collect opinions from various sources such as companions and are influenced by what we think about that particular in various types of the decision-making process. Assessment is a review concerning some article or topic etc. On the development and appearance of web 2.0, various websites have expanded complex such as Twitter, LinkedIn, Facebook, etc. This development increases the number of users and their content on these platforms. It examines the user's feelings for some specific topic or product. It is monstrously helpful in web-based data checking since it permits to the acquisition of knowledge for popular assessment on some specific subjects. This study has handled a complete outline of the Lexicon method and its implementation updation from time to time. Numerous proposed improvements, various challenges faced in sentiment analysis, and different Sentiment analysis applications are introduced and explored in this study.

Keywords: Lexicon-based techniques, Opinion mining, Sentiment analysis, Sentiment classification, Social networking.
SAMRIDDHI : A Journal of Physical Sciences, Engineering and Technology (2022); DOI: 10.18090/samriddhi.v14i01.14

Introduction

SA utilized[1] NLP, text mining, and various computational methods to extract reviews from a given review. This method is applicable in a large number of fields such as social media, business analysis, customer information, etc. The various social media platforms and online media are becoming more active, increasing online users' data. Due to the huge data available on the internet as social media, it is very tough to find exact trends about the particular item/product on the various e-commerce platforms. It requires some automated systems or applications that can easily find current trends of sentiments on some valuable data. Sentiment analysis, likewise named opinion mining, is the mining of assessments of people, their evaluations, and emotions toward specific articles, realities, and characteristics. Most sentiment work is based on the two main strategies: the traditional approach and the semantic approach. In the traditional technique, featured words are used, while contextual words are used for the sentiment analysis process in the semantic approach. The effect of phonetics develops like conditionals; conjunctions and intensifiers on sentence-level sentiment analysis have been examined. The online review[2] aims to analyze and examine the review's sentiment score. The sentiment analysis is divided into three main categories, document level,[3] we investigate structured models for document-level sentiment classification. When predicting the sentiment of a subjective document (e.g., as positive or negative sentence level,[4] and word level.[5] The popular technique[6] was found suitable

Corresponding Author: Surendra Kumar, Research Scholar, Department of Computer Science and Engineering, Dr.K.N.Modi University, Newai, Rajasthan, Newai, India, e-mail: surendra.ftp@gmail.com

for sentiment analysis. They have furnished a perspective for sentiment analysis techniques and challenges facing in research on today's tren.[7] suggest a logical model analysis that the online audit information finds the actual facts about the item; outcomes showed that 53% of surveys had a bimodal. Non-typical conveyance and those audits can't be assessed with the normal score. In this manner, a model was characterized to clarify when the mean can fill in as the legitimate portrayal of an item's genuine quality.

The dataset which is utilized a significant piece of opinion mining. The primary source of information/data is from the user's item review. The review given by the users gives understanding into an item quality and response from the users that may be utilized to implement insignificant for commercial actions. Sentiments are mostly used to take concern about the product available on the eCommerce site and other general issues in social networking sites due to the development of online produced data.

Twitter is a microblogging platform in which a large number of users may trade their sentiments. The information on the Twitter platform is a prudent and compelling approach to uncovering general sentiment. There is some exploration of tasks that concentrate on the estimation of the relativeness of the online reviews opinion and real-life opinion (review of a new product, stock market prediction,[8] poll opinion.[9] This is accounted for that occasions, in actuality, genuinely have a huge and quick impact on the public sentiment on Twitter. In light of such relationships, some different tasks [10] utilized the opinion in websites and tweets to anticipate film revenue[11] and various decisions. Pang Lee[12] and Liu[13] introduced the various difficulties and applications in the area of sentiment analysis also introduced the strategies to tackle the challenges in SA.

PROCESS OF THE SENTIMENTAL ANALYSIS

Review Extraction

The information available is available on the article summary pages of various e-commerce sites. The information sites may be extracted from various social networking sites utilizing the API. The API is a kind of application that is used to extract the data from the given particular souces. For instance, on account of Twitter, we can extract information utilizing Twitter API. From the given link "https://developer.twitter.com/en/docs ."By making registration and getting approval from Twitter, we can get data in CSV file format, which can be used for processing.

Pre-processing

Pre-preparing is a way toward cleansing collected information preparing the content for the classification process. The pre-processing steps are described as in Figure 1. E-content

Figure 1: SA process

writing normally contains not use data and pointless parts like labels and contents. Pre-processing technique decreases reduce the unnecessary information, which assists with improving the performance of the classifier. Pre-preparing additionally accelerates the classification, consequently helping continuously in sentiment analysis. Haddi, Emma, Xiaohui Liu, and Yong Shi.[14] depicted that proper content pre-handling having information changes, sifting and data can be fundamentally enhanced classification process.

For instance, for a system that analyses sentiment analysis of Twitter data for twitter in the English language, the following pre-processing process:

*   Removing words not in English- non-English tweets should be removed as we perform SA of the English tweets only.
*   Removal URLs, numbers, punctuations, no-meaning words, references, the special character used, and hashtags. The removal of the URLs, no meaning words, any reference, any special character used, white space, hashtags, tabs, etc., reduces the data's noises, which is necessary to increase the performance. The "RT" value should be null to get desired output.
*   Translation of the Slang word–Slang words are the words that are used in the sentences but don't have an exact meaning, and these words are to be replaced by some meaningful words which are relevant in that sentence. For example." He invited all batch mates to his birthday. He is too extra" here, extra is a slang word is which is to be replaced.
*   Removal of Unnecessary letters from words- Some reviews have some extra letters(same letters) in a word. The repeating letters are to be removed so that it occurs only once and should not be available in the Lexicon. As word "Enjoyyyyyyyyyyyyy" is replaced to "Enjoy."
*   Stemming – The suffixes are to be removed from the words so that the words appear in their base word as "walking," "walk," "walked," to be substituted by "walk."

Feature Identification

The data gathered subsequently got is as words. These words are known as tokens. There is no ambiguity in the data. The main step in text classification is choosing the terms to be utilized as properties in the training data.[15] Choosing the feature word has benefits like boosting up the training process due to reduced words and increasing the classification accuracy by removing noisy data. The intention is to discover the sentiment expressions that best fit, aiming word in the given data. This step is a fundamental step in various opinion mining applications.[16] proposed a method to recognize opinion words that do not include the seed words.

Classification

Classification is methods can be separated in two types of applications. Specifically, the ML and Lexicon based. ML techniques depend on the training of the model, for the
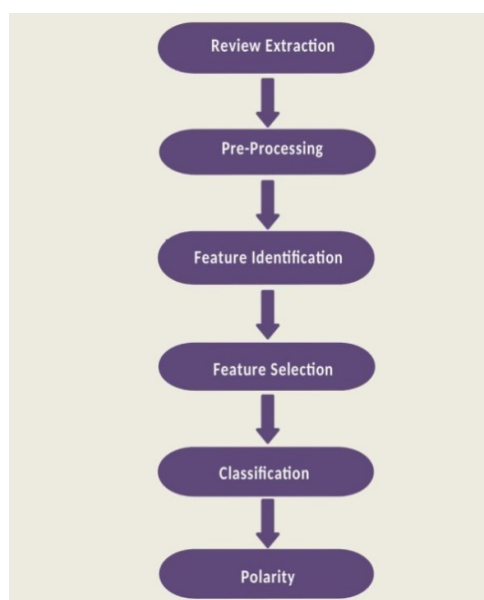
most part, classification process, chosen featured data for a particular aim, afterward test on different data*set al*though, it can identify the correct features and classificify them. Pang, Bo, Lillian Lee, and Shivakumar Vaithyanathan[12] have added to classification utilizing feature attributes.

Polarity

Polarity is also called sentiment orientation. It determines the orientation of the given sentiment. In the Lexicon-based method, the polarity is calculated of each word and then the overall polarity of the sentiment using dictionary.[17] The Lexicon approach gives high accuracy.[18]

The Ml techniques are very interesting techniques for researchers because of the nature of their ability and accuracy. The ML techniques have two types one is supervised, and another is unsupervised. The sentiment analysis uses a traditional learning-based model, which is incorporated in two ways of learning, either supervised or unsupervised. In the supervised learning technique, the raw data is divided into two basic parts: training and test sets. The training set is utilized with the machine trained enough

to classify new samples coming from the test set. The most powerful techniques for sentiment analysis are support vector Machine, Neural Network, and LDA-based model.

LEXICON-BASED APPROACH

Lexicon Based Approach determines the polarity of the given text based on the polarity of any sentence. The various methods are described as in Figure 2. This methodology depends on estimation vocabularies that are generally accessible. The vocabularies map words to supposition scores (negative, positive or impartial) that are utilized to anticipate the assumption of the content. The slant likelihood of words is normally prepared from semantic corpora.[19] A vocabulary-based methodology is an unsupervised approach; a solo methodology in this way has an advantage that does not require labeling of the dataset. Initial, a group of words and their separate semantic direction, is to be gathered in a word reference. Second, the extraction of an adjective is to be completed and comment those with Semantic direction esteems, using word reference scores. At last, these semantic direction scores are joined into a solitary score.[20]
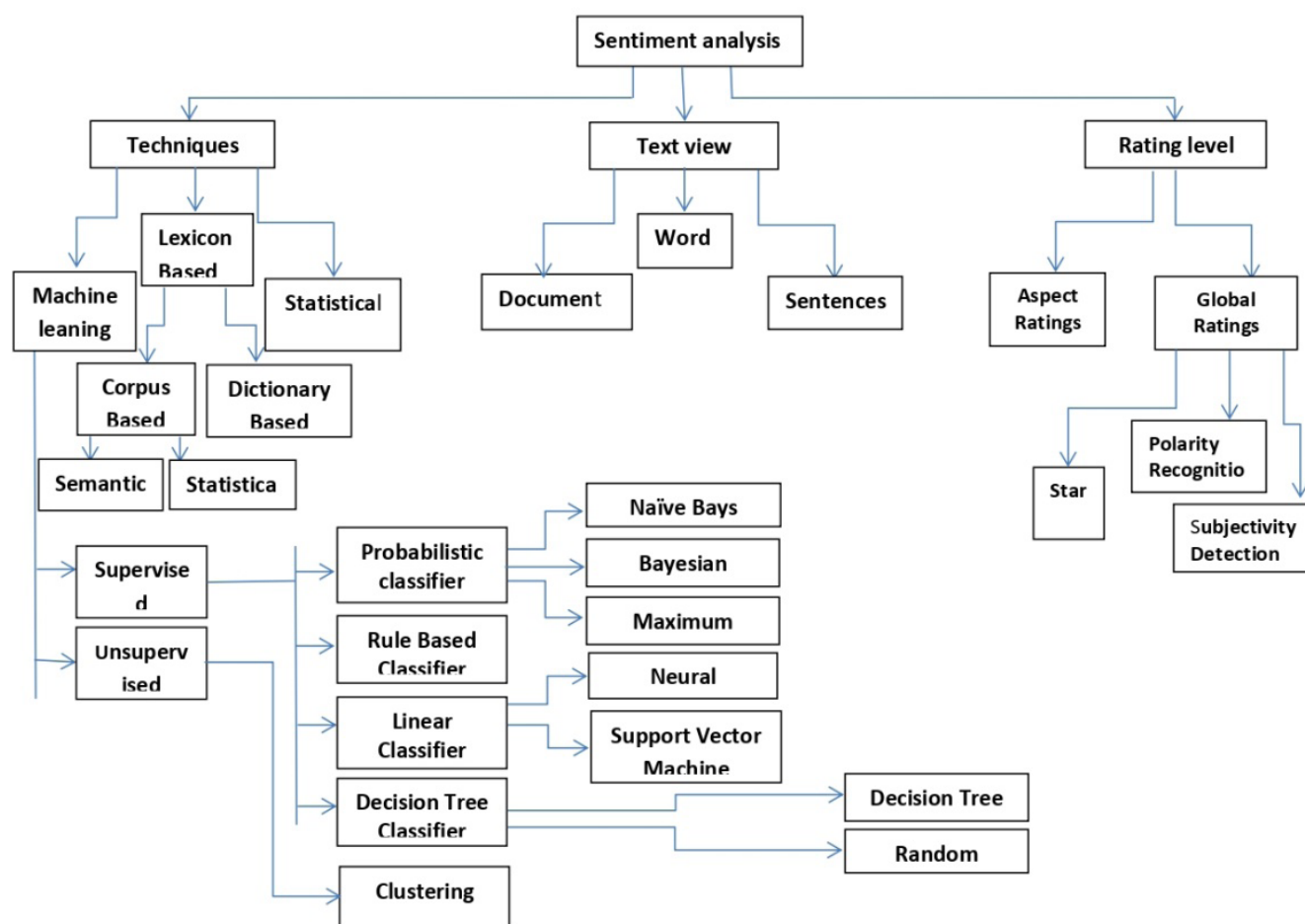
SENTIMENT ANALYSIS TECHNIQUES

Figure 2: Techniques of SA

This method primarily relies on getting predilection of any text files and incorporates calculating predilection for a report from the given words or terms inside the report.[21] The lexicon-based technique is mainly based on the text. The procedures are created manually or maybe automatically[22] to use seed phrases to amplify the listing of phrases. Most of the lexicon-primarily based total studies have centered on the usage of adjectives as signs of the semantic orientation of the textual content. The listing of adjective words and matching words values are collected right into this created or automatic dictionary. According to preceding research, the adjective is a true sign of sentiment orientation.[23] In the textual content, all adjectives are fetched from the text files and marked SO value using the dictionary created. However, even though an out-of-the-manner adjective can also display biased opinion, there can also be an unsatisfactory context to decide semantic orientation. As discussed via way of means of P. Turney,[21] the adjective "brutal," insane" can also additionally have a poor polarity in poor assessment, in a word consisting of "brutal and insane breed of dog," although it is having a good polarity of review in a film. The words as "brutal and insane action sequence." In this case, the rules of extraction extract many continuous phrases. It extracted the adjectives, words, and phrases that are given in the context.

The lexicon-based technique has disadvantages like the word "high" in the context of "protein" has a positive polarity, and "high" in view of "alcohol" has a negative polarity. This technique can be categorized into following ways: Some of the Lexicon based approach, hybrid approach and ensembles approach, papers are compared as in Tables 1-3 respectively.

Dictionary-based

This method uses a dictionary that is assigned a number called PoS tagging, and each word is assigned a score by its polarity.[24] The dictionary-Based method entails the usage of a repository of words called a dictionary. It is incorporated with a group of synonyms words and antonyms of the phrases. Thus, an easy method is to apply some sentiment phrases to assist primarily based totally on this dictionary. This technique works on a group of some sentiment phrases collected and acknowledged with a manual process providing the polarity such as poor, good, better, etc. This is a very basic process. The set of rules then expand this group via way of means of looking in any online to be had a dictionary for his or her synonyms and antonyms. Newfound phrases can bring the source words. The system iteratively continues, including phrases or words, till more suitable new phrases is discovered. A manual review may be applied to smooth up the process.

Corpus-based

The Corpus-primarily based totally method allows resolving the trouble of locating opinion phrases with context-particular references. These techniques rely on a syntactic approach, which arises collectively unitedly with seed listing words of the sentiment phrases to locate different sentiment phrases in large amounts of data. This method is classified as:

Statistical Approach

Although phrase seems irregularly among good texts given, its polarity will also be good. If there is the occurrence of more negative words, then the polarity may be taken into consideration as poor polarity. If there are occurrences of same words, then it may be an impartial phrase. Seed sentiment phrases may be discovered the usage of statistical strategies. The most natation of art methods is primarily based totally on the statement that comparable opinion phrases, in general, seem collectively in a corpus. Thus, if two phrases seem collectively often in the identical context, then there may be an excessive case as they may assign the same polarity. Theo finds that the polarity of an obscure phrase may be decided by calculating the relative frequency of coincidence with some different phrase. This may be achieved the usage of Pointwise Mutual Information (PMI) as with inside the example suggested[21] via way of means of, SO of a given the word is calculated via way of means of evaluating its similarity to a fantastic phrase ("Awesome"), it is similar as the poor phrase ("Awful"). As a word is assigned, a numeric number called rating is a way of collecting the common records among the given word. The good acknowledge phrase "Awesome" eliminates the common records among the word, and the poor acknowledge phrase. "Awful." A (POS) method classifies the textual content via way of means of extracting the bigrams. PMI is then calculated via way of means of the usage of the polarity rating for every context.

Semantic Approach

Semantic approach precept delegates comparable opinion value to semantically near phrases. This Semantically near word may acquire via way of means of getting the listing of sentiment phrases and step by step increasing the preliminary set with synonyms and antonyms after which figuring out the sentiment polarity for an unknown phrase via way of means of the relative reliability of negative as well as positive synonym of the phrase.

APPLICATION OF SENTIMENT ANALYSIS

E-commerce- The sentiment analysis is mostly applied in many e-commerce organizations. The enterprises' websites allow their users to give rates or review their own experiences about purchasing and quality of the product or any services taken from that organization. The company provides an overall rating for that product, and people can take decisions on behalf of the reviews given by the online users.[37] Normally we buy some product online if it has a positive sentiment of review given by the users.

The sentiment analysis is an important aspect from the company's point of view. Based on product review they can improve the quality of the product which increases the profit. The customers who write an unsatisfactory review analyze and try to rectify the problems.

The SA is applied in the recombination system in e-commerce sites. Normally any product recommended to a

REVIEW LITERATURE

Table 1: Lexicon-based methods

Study	Xia et al. [25]	Azzouza, Noureddine, et al. [26]	Paltoglou and Thelwall [27]	Masood et al.[28]	Asghar et al. [29]
Method	Unsupervised method	Unsupervised method	Unsupervised method (lexicon-based)	Unsupervised method (lexicon-based)	Lexicon-enhanced-Rule-based
Algorithm	Focused on slang words	Lexicon Based	Emotional Lexicon	Lexicon and dictionaries	Rule-based classifier
Feature	Unigrams	POS features	Unigrams	own datasets	An improved feature weight technique. Emoticon
Dataset	STS and OMD datasets	SemEval-13 to 16	Digg, MySpace, and Twitter datasets	own datasets	Three review datasets
Outcome	72% accuracy achieved for STS and 69.2% for OMD	55.9% accuracy achieved for SemEval-13,53.6& for SemEval-16 & 50% achieved by SSA-UO and 53.9 gained by GTI	Achieved F1-measure value 76.2 for Digg, 80.6 for myspace, and 86.5 for Twitter.	Gained accuracy by 92% and 87% for binary classification and 8 multi-class classifications.	For the IInd dataset, get F1score by79.5%.

Table 2: Various hybrid methods

Study	Balage Filho and Pardo [30]	Khan et al. [31]	Zainuddin et al.[32]	Asghar et al.[33]
Methods	Hybrid Approach	Hybrid Approach	Hybrid Approach	Hybrid Approach
Algorithm	Support vector machine, rule-based method, and Senti-Strength for lexicon approach	SentiWordNet Classifier, Enhanced Emoticon Classifier, and Improved Polarity Classifier	Principal component analysis and the SVM classifier.	Slang classifier, Emoticon classifier, and improved domain-specific classifier.
Feature	BOW	SentiWordNet Emoticons, sentiment words	Stanford dependency parser Association rule mining and part of speech	SentiWordNet
Dataset	SemEval-2013 Task datasets	Own datasets	STS, HCTS, and STC datasets	Own datasets
Outcome	This technique achieved F-score of 56.3% and, compared to SVM, achieved 49.9%	85.7% accuracy, 85.3% precision, and 82.2 recall were achieved.	This technique applied with the STS achieved 76.55% accuracy, with HCTS 71.62%, and with STC 74.24%	This technique achieved F-score by 88%.

Table 3: Ensemble techniques in opinion mining

Study	Lin, J and Kolcz, A. [34]	Hagen, M., Potthast, M., Büchner, M., & Stein, B. [35]	Chalothom, T, and Ellman, J. [36]
Methods	Ensemble	Ensemble	Ensemble
Algorithms	Logistic regression classifier	TeamX NRCcanada, GU-MLT-LT, and KLUE	Stacking, SVM, Naïve Bayes, and Senti-Strength
Features	Hashed byte 4-grams	Stemming, punctuation-marks, emoticons dataset used, word-length, negation, parts of speech, and polarity of the sentence	Sentiment lexicons and BOW features
Datasets	Large-scale datasets	SemEval-2013 training	SemEval-2013
Outcomes	Achieved an accuracy of 81% using 21 Classifiers over 100 million instances.	Achieved F-score by 64.84 %.	Achieved F-score by 86.05%.

user in the e-commerce sites which have a positive sentiment. The recommendation system does not recommend any negative review product.

In the field of marketing, the voice of customers and the voice of market techniques are helping to be aware of the market's current scenario that is totally based on SA. The companies who find early become the leading company because there is lots of completion between companies of the same field.

CHALLENGES IN SENTIMENT ANALYSIS

The various languages are there, and we need to design the application in that language, so it is a very challenging task to make content available in a particular language and find the sentiment. Multilingualism is a major concern in the field of SA.[38] Many researchers are working on transforming the model from one domain to another domain. The scalability is required more for an application because every day, each minute, lots of data is generated, and applications need to be good analyzers and process these data. Robustness is also required and it's a challenge. Every human language can be understood by everyone as well as by machines. A new word, text, of linguistic grammar, is not possible make learn to the system. It is very hard to find the given reviews are genuine or not. Many companies are using the exercise of fake reviews to promote the product.

REFERENCES

[1] B. Agarwal, N. Mittal, P. Bansal, and S. Garg, "Sentiment analysis using common-sense and context information," *Comput. Intell. Neurosci.*, vol. 2015, pp. 1–9, 2015, DOI: 10.1155/2015/715730.

[2] L. Peng, G. Cui, M. Zhuang, and C. Li, "What do seller manipulations of online product reviews mean to consumers?," *account. Finance.*, vol. 24, no. 1, pp. 75–75, 2009.

[3] A. Yessenalina, Y. Yue, and C. Cardie, "Multi-level structured models for document-level sentiment classification," *EMNLP 2010 - Conf. Empir. Methods Nat. Lang. Process. Proc. Conf.*, no. October, pp. 1046–1056, 2010.

[4] N. Farra, E. Challita, R. A. Assi, and H. Hajj, "Sentence-level and document-level sentiment mining for Arabic texts," *Proc. - IEEE Int. Conf. Data Mining, ICDM*, pp. 1114–1119, 2010, DOI: 10.1109/ICDMW.2010.95.

[5] N. Engonopoulos, A. Lazaridou, G. Paliouras, and K. Chandrinos, "ELS: A word-level method for entity-level sentiment analysis," *ACM Int. Conf. Proceeding Ser.*, 2011, DOI: 10.1145/1988688.1988703.

[6] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams Eng. J.*, vol. 5, no. 4, pp. 1093–1113, 2014, doi: 10.1016/j.asej.2014.04.011.

[7] M. Hu and B. Liu, "Opinion extraction and summarization on the web," *AAAI*, pp. 1–4, 2006, [Online]. Available: http://www.aaai.org/Papers/AAAI/2006/AAAI06-265.pdf.

[8] J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market," *J. Comput. Sci.*, vol. 2, no. 1, pp. 1–8, 2011, DOI: 10.1016/j.jocs.2010.12.007.

[9] Marvell Solutions, "88E111 Datasheet," *Fourth {International} {AAAI} {Conference} on {Weblogs} and {Social} {Media}*, 2010, [Online]. Available: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/view/1536.

[10] G. Mishne and N. Glance, "Predicting movie sales from blogger sentiment," *AAAI Spring Symp. - Tech. Rep.*, vol. SS-06-03, pp. 155–158, 2006.

[11] I. Tumasjan, A., Sprenger, T., Sandner, P., & Welpe, "Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment," *Proc. Fourth Int. AAAI Conf. Weblogs Soc. Media*, vol. 4, no. 1, pp. 178–185, 2010.

[12] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? Sentiment Classification using Machine Learning Techniques," 2002, DOI: http://dx.doi.org/10.3115/1118693.1118704.

[13] B. Liu, *Sentiment Analysis and Opinion Mining*, no. May. 2012.

[14] E. Haddi, X. Liu, and Y. Shi, "The role of text pre-processing in sentiment analysis," *Procedia Comput. Sci.*, vol. 17, pp. 26–32, 2013, doi: 10.1016/j.procs.2013.05.005.

[15] M. Abulaish, M. N. Doja, and T. Ahmad, "LNCS 5909 - Feature and Opinion Mining for Customer Review Summarization," pp. 219–224, 2009.

[16] H. Yu, Z. H. Deng, and S. Li, "Identifying sentiment words using an optimization-based model without seed words," *ACL 2013 - 51st Annu. Meet. Assoc. Comput. Linguist. Proc. Conf.*, vol. 2, pp. 855–859, 2013.

[17] C. Bhadane, H. Dalal, and H. Doshi, "Sentiment analysis: Measuring opinions," *Procedia Comput. Sci.*, vol. 45, no. C, pp. 808–814, 2015, DOI: 10.1016/j.procs.2015.03.159.

[18] M. Annett and G. Kondrak, "A comparison of sentiment analysis techniques: Polarizing movie blogs," *Lect. Notes Comput. Sci. (including Subsea. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 5032 LNAI, no. Figure 1, pp. 25–35, 2008, DOI: 10.1007/978-3-540-68825-9_3.

[19] D. Rao and D. Ravichandran, "Semi-supervised polarity lexicon induction," *EACL 2009 - 12th Conf. Eur. Chapter Assoc. Comput. Linguist. Proc.*, no. April, pp. 675–682, 2009, DOI: 10.3115/1609067.1609142.

[20] A. Jurek, M. D. Mulvenna, and Y. Bi, "Improved lexicon-based sentiment analysis for social media analytics," *Secure. Inform.*, vol. 4, no. 1, 2015, doi: 10.1186/s13388-015-0024-x.

[21] P. D. Turney, "Turney-Acl02-Final," p. 8, 2006.

[22] A. Goyal and H. Daumé III, "Generating Semantic Orientation Lexicon Using Large Data and Thesaurus," *Proc. 2nd Work. Comput. Approaches to Subj. Sentiment. Anal.*, pp. 37–43, 2011.

[23] V. Hatzivassiloglou and K. R. McKeown, "Predicting the semantic orientation of adjectives," pp. 174–181, 1997, DOI: 10.3115/979617.979640.

[24] S. Almatarneh and P. Gamallo, "A lexicon-based method to search for extreme opinions," *PLoS One*, vol. 13, no. 5, pp. 1–19, 2018, DOI: 10.1371/journal.pone.0197816.

[25] X. Hu, J. Tang, H. Gao, and H. Liu, "Unsupervised sentiment analysis with emotional signals," *WWW 2013 - Proc. 22nd Int. Conf. World Wide Web*, pp. 607–617, 2013, DOI: 10.1145/2488388.2488442.

[26] N. Azzouza, K. Akli-Astouati, A. Oussalah, and S. Ait Bachir, "A Real-Time Twitter Sentiment Analysis using an unsupervised Method," *ACM Int. Conf. Proceeding Ser.*, vol. Part F1294, 2017, DOI: 10.1145/3102254.3102282.

[27] G. Paltoglou and M. Thelwall, "Twitter, MySpace, Digg: Unsupervised sentiment analysis in social media," *ACM Trans. Intell. Syst. Technol.*, vol. 3, no. 4, 2012, DOI: 10.1145/2337542.2337551.

[28] M. A. Masood, R. A. Abbasi, and N. Wee Keong, "Context-Aware Sliding Window for Sentiment Classification," *IEEE Access*, vol. 8, pp. 4870–4884, 2020, DOI: 10.1109/ACCESS.2019.2963586.

[29] M. Z. Asghar, A. Khan, S. Ahmad, M. Qasim, and I. A. Khan, "Lexicon-enhanced sentiment analysis framework using rule-based classification scheme," *PLoS One*, vol. 12, no. 2, pp. 1–22, 2017, DOI: 10.1371/journal.pone.0171649.

[30] P. P. Balage Filho and T. A. S. Pardo, "NILC USP: A hybrid system for sentiment analysis in twitter messages," **SEM 2013 - 2nd Jt. Conf. Lex. Comput. Semant.*, vol. 2, no. SemEval, pp. 568–572, 2013.

[31] F. H. Khan, S. Bashir, and U. Qamar, "TOM: Twitter opinion mining framework using hybrid classification scheme," *Decis. Support Syst.*, vol. 57, no. 1, pp. 245–257, 2014, DOI: 10.1016/j.dss.2013.09.004.

[32] N. Zainuddin, A. Selamat, and R. Ibrahim, "Hybrid sentiment classification on twitter aspect-based sentiment analysis," *Appl. Intell.*, vol. 48, no. 5, pp. 1218–1232, 2018, DOI: 10.1007/s10489-017-1098-6.

[33] M. Z. Asghar, F. M. Kundi, S. Ahmad, A. Khan, and F. Khan, "T-SAF: Twitter sentiment analysis framework using a hybrid classification scheme," *Expert Syst.*, vol. 35, no. 1, pp. 1–19, 2018, DOI: 10.1111/exsy.12233.

[34] J. Lin and A. Kolcz, "Large-scale machine learning at Twitter," *Proc. ACM SIGMOD Int. Conf. Manag. Data*, pp. 793–804, 2012, DOI: 10.1145/2213836.2213958.

[35] M. Hagen, M. Potthast, M. Büchner, and B. Stein, "Webis: An Ensemble for Twitter Sentiment Detection," no. SemEval, pp. 582–589, 2015, doi: 10.18653/v1/s15-2097.

[36] J. Chalothom, T., & Ellman, "Simple approaches of sentiment analysis via ensemble learning," *Inf. Sci. Appl.*, pp. 631–639, 2015.

[37] S. Tan *et al.*, "Interpreting the public sentiment variations on Twitter," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 5, pp. 1158–1170, 2014, DOI: 10.1109/TKDE.2013.116.

[38] Y. Seki, D. K. Evans, L.-W. Ku, L. Sun, H.-H. Chen, and N. Kando, "Overview of multilingual opinion analysis task at NTCIR-7," *Proc. NTCIR7 NII Tokyo*, vol. 1, pp. 185–203, 2008, [Online]. Available: http://nlg3.csie.ntu.edu.tw/conference_papers/nntcir2008b.pdf.