Sentiment Analysis Techniques –Survey

Nobogh Husssein Baqer (✉)
Department of Computer Science, College of Education, Mustansiriyah University, Iraq.
nobogh.hussein@uomustansiriyah.edu.iq

Zuhair Hussein Ali
Department of Computer Science, College of Education, Mustansiriyah University, Iraq.
zuhair72h@uomustansiriyah.edu.iq

Abstract: Online social media platforms allow people to express their ideas and opinions about anything as products, education, or politics, this produce a huge data requires appropriate tools to analyze it accurately, and that's why Sentiment Analysis (SA) has become more used in recent years. This paper review the SA and classification methods of machine learning especially supervised techniques like Support Vector Machine (SVM), Neural Networks (NN), Naïve Bayes (NB), Rough Set Theory (RST), Random Forest (RF), Decision Trees (DT). The other technology of SA is lexicon techniques and hybrid methods (which include the lexicon and machine learning techniques). Machine learning consist of two group supervised (Classification and Regression) and unsupervised (clustering and association). The SA Affected from one system to another based on its polarity as "positive" or "negative", and the language of data that processed.

Keywords- Lexicon techniques, Machine Learning techniques, sentiment classification.

1. Introduction

SA is a method utilized to analyze opinions on a particular issue, or products often extracted from social media platforms, where it considers as communication environment help people to express their opinion. SA classify the opinions based on its polarity depending on 3 classes "positive", "negative", or "neutral", Natural sentences are those that do not represent a negative or positive sentiment[1].

According to the continuously evolution of technology, SA become more trending in research area, where various types of Classifiers are used for opinion mining SA divided into three groups' machine learning techniques, lexicon techniques and hybrid methods [2]. Machine learning algorithm particularly used in SA because it perform better than other methods based on the measurement of standard evaluation metrics (Precision, Recall, F-measure, Accuracy) and the amount of time saving. Machine learning requires labelled dataset for training and includes a set of classifiers; decision tree, rule-based, probabilistic (Includes: Naïve Bayes, Maximum Entropy, and Bayesian Networks), and linear-based (Includes: SVM and neural networks). Lexicon-driven methods are based on dictionary and other specific pre-tagged types in parts of textual. There are many methods of lexicon based algorithms, some of them give the ability to sets and detect the sentiments in 5 categories as quite positive, positive, neutral, negative and quite negative, and it includes two approaches dictionary based and corpus based (which includes statistical and semantic). The last method is the hybrid method which merge lexicon and machine learning techniques. Sentiment lexicons usually plays a main role for these strategies[3][4].

This paper will explain following questions; what is SA? And why we need to have classification techniques, and what is those technique? The objective of this paper to produce a taxonomy of SA techniques help in choosing analytical techniques for future researches, where SA applied on Media, education, e-commerce and others fields[5]. The research order is; the related work is provide in section 2, where section 3 for the research method, and Section 4 provides the conclusion.

2. Related work

There is a lot of research on SA techniques. Below will display some of the previous works on this subject:

V.Elango and G.Narayanan in 2011, utilize hotel reviews dataset to classify reviews as a "positive" or "negative" than analyzing the customer sentiment. To extract the frequent words from the reviews used Term Frequency-Inverse Document Frequency (TF-IDF). And utilizing different classification techniques to classify the review and get the accuracy, which are; Support vector machine (SVM) 75.29%, Naïve Bayes Multinomial 79.12%, and Naïve Bayes Bernoulli 78.86%, when increasing the training data, the accuracy rate increases. Based on excremental results, it was found that Machine Learning techniques can definitely outperform human-produced SA baselines[6].

Jadav and Vaghela in 2016, pre-processed the dataset to convert it into a structured model by using feature selection and semantic analysis, then use lexicon approach to convert the review into numerical value. Remove stop word, stemming, and calculate sentiment score by using SentiWordNet dictionary this done in pre-processing portion. To classify opinion for positive or negative applied classification algorithm by using SVM algorithm to classify reviews. After compare performance of all classifier found that optimized SVM gives better result than NB and SVM according to their accuracy[7].

V. Umarani and others in 2021, use restaurant dataset to analyzing their opinion as "positive" or "negative". To enhance the efficiency of the classification task, the initial step involves preprocessing to minimize the number of features. Using the machine learning techniques to classify the sentiments (Multinomial Naive Bayes, Bernoulli Naive Bayes, Logistic Regression, Random Forest, KNN, SVC, and Decision Tree) and the deep learning techniques (LSTM and CNN). In machine learning using bag of model methods (TFID) and in deep learning used word embedding to convert text into vector. Then using confusion matrix, AUC score, and time taken for training to evaluate the performance of classifiers. While accuracy score for each classifiers tested through k-fold cross validation technique. Based on the experimental results, it has been observed that neural network-based learning exhibits superior training accuracy, albeit with a longer execution time in comparison to machine learning classifiers[8].

A. Punitha and others in 2021, combines data from website and recognize hard terms sentences by transformation and sentiment intensity is predicted in efficient manner. This done by develop application to fetch the real time texted data of the user which is feed as input to the trained model, this help overcomes the drawback of manual feed and regular feeding of data. To predict the mental discomfort of a person uses SVM with accuracy of 95%. The real-time detection of messages helps to analyze the mental discomfort of a person at an earlier stage[9].

Nadia et. al. in 2021, proposed model that assess companies e-tourism by using Iraqi dialect reviews gathered from Facebook. This reviews are analyzed through text mining methods for sentiment classification. The sentiment words generated are classified as 'positive', 'negative' and 'neutral' posts by uses Rough Set Theory, K-Nearest Neighbor and Naïve Bayes techniques. In the experimental results 71 Iraqi tourism companies were tested, 28% from these companies have pretty good assessment, 26% from these companies have a good assessment, 31% from these companies have middle assessment, 4% from it have accepted assessment and 11% have very bad assessment. These results assisted the companies to enhance their work and programs responding enough and fast to customer demands[10].

Saman Zahid in 2021, implement SA on user reviews using 3 types of classifiers, which are; Naive Bayes, Random Forest, and Support Vector Machine. After calculate the confusion matrix of each one, found that SVM performs better than other classifiers with accuracy about 81.6%, and 66.5% F1-score. The reviews are classified as positive, negative or natural labels[11].

Abeer et. al. in 2022, develop a method of SA to classify the beauty centers in Iraq into healthy and unhealthy categories. Researcher's usage comments of beauty centers' Facebook to implement the assessment. The methodologies comprise the two methods of lexicon-based and machine-learning-based. Three machine learning mechanisms had been implement; rough set theory, naïve bayes, and k-nearest neighbors. Its notes that rough set theory is better comparison with other two. Rough set theory reach to 95.2%, while Naïve Bayes reach to 87.5% and k-nearest neighbors reach to 78% [12].

3. Research method

This section explains the procedure for conducting research that include the SA definition and the different techniques used in it, and the pros and cons of these techniques. There are several surveys has described SA techniques, this research will perform a survey of the available SA techniques and taxonomy.

3.1. Sentiment Analysis

SA (or opinion mining) utilizes a natural language processing (NLP) to reproduce a type of knowledge from a big amount of data. It scans opinions, comments, emotions, views, attitudes the writer of the message or the text. It mining the writer's emotions in the form of subjectivity as "objective" and "subjective", polarity as "positive", "negative", and "neutral", and emotions like "happy", "angry", "surprised", "sad", and "jealous"[13].

SA is an intellectual process of mining the user's emotions. The tremendous development of the Internet has led to its frequent use by users to express their opinions in all fields. These reviews found in various forms like social media, blogs or websites forum. And found reviews information are a powerful tool, it's useful for understanding and planning processes for users for reproduce the necessary information. To accomplish this task, in past years several methods have come to the lights [14].

3.2. Sentiment classification techniques

The main objective of the classification is to introduce view for simple understanding. Before classification step, cleaning data in the preprocessing stages and feature extraction were implemented. Several deep learning methods were used in the training stage.[5] There are many algorithm can be used that help in the knowledge of positive or negative emotions and produces result with vary accuracy. Basically, 2 kinds of Sentiment classification algorithms exist: Machine Learning and Lexicon-based approaches[15].

In Taxonomy tree architecture display categorize of SA methods depending on their application in SA as shown in Figure 1. This techniques cover approximately all areas of application of SA of education, movie, marketing, etc. Depending on Figure 1 it noticed that SA techniques have two section: Machine learning and lexicon-based methods. The machine learning requires too much of labelled data to give best accuracy result, and its characterized as unsupervised and supervised algorithms that divide into four part; decision tree classifiers, rule-based classifiers, linear classifiers (include SVM and Neural networks), and probabilistic classifiers (include Naïve Bayes, Bayesian network and maximum entropy). The lexicon-based approaches don't need labelled data because its unsupervised, and its split to two portion, dictionary and corpus based approaches (include statistical and semantic)[16].

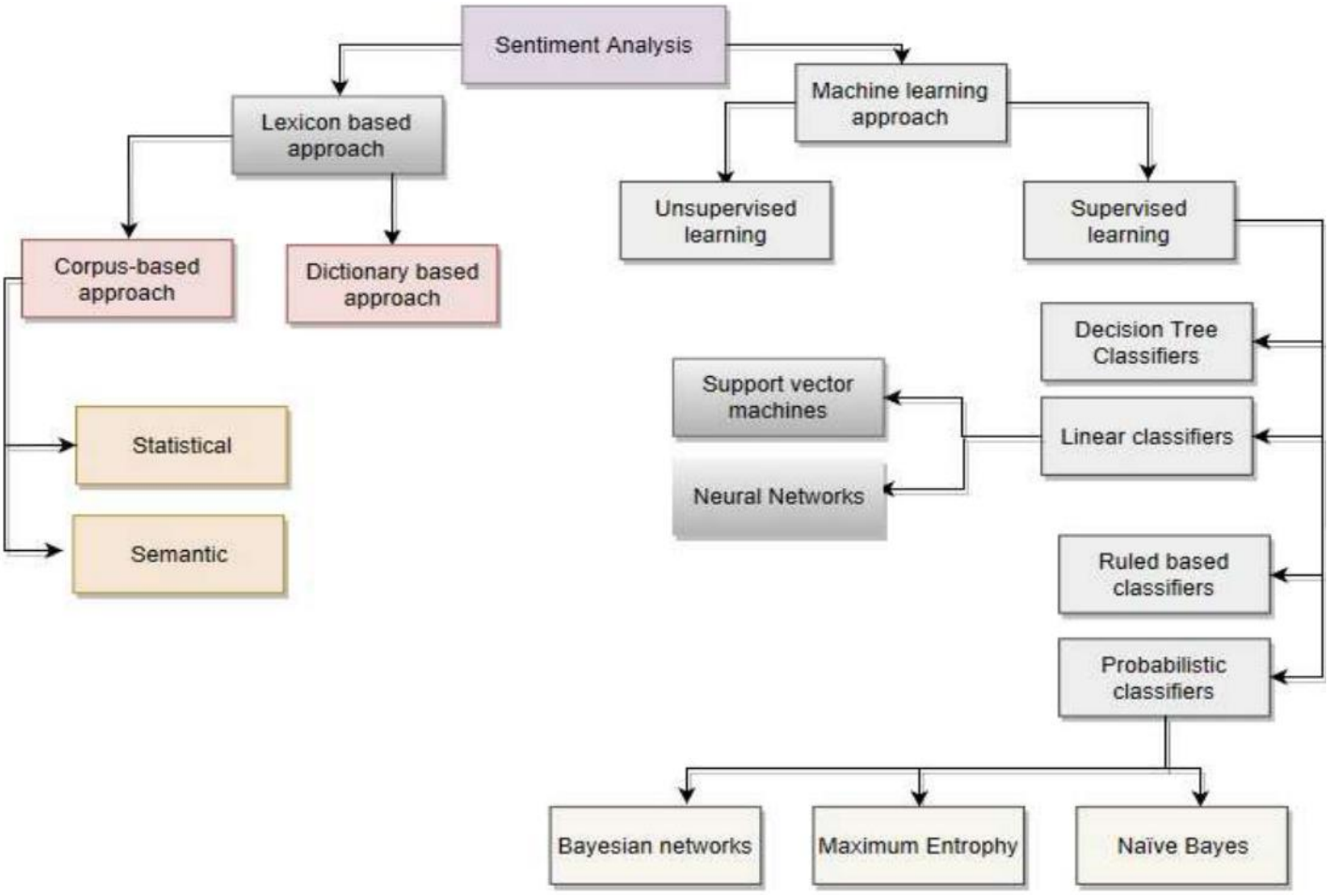

Fig 1. Sentiment Analysis Classification Techniques[5].

3.2.1. Machine learning approach

Machine learning is a critical component of the computer science development, by using statistical methods, trained algorithms for classification or prediction, this lead to decision making within applications and businesses[17]. Machine learning for Sentiment classification is composed of two steps: First, the extracted features are collected in vectors and second, using the classification algorithms in training this vectors[1]. This method have some advantages and disadvantages that depending on the specific application, the quality of data, and the choice of machine learning algorithms. Where the advantages can be as following:

1. The efficiency: Machine learning enables the automation of complex tasks, reducing the need for manual intervention and improving efficiency.
2. Complex Data: Machine learning algorithms can effectively handle large and complex datasets, including unstructured data such as text, images, and videos.
3. Pattern Recognition: Machine learning algorithms excel at identifying patterns and relationships within data that may not be apparent to humans, leading to valuable insights and predictions.

4. Adaptability: Machine learning models can adapt and improve over time as they are exposed to more data, allowing for continuous learning and better performance.

While the disadvantages of machine learning are:

1. Data Dependence: Machine learning models heavily rely on the quality and quantity of training data. Insufficient or biased data can lead to inaccurate or biased results.
2. Over-fitting: Machine learning models can over-fit the training data, meaning they become too specific to the training set and fail to generalize well to unseen data.
3. Computational Requirements: Complex machine learning models often require significant computational resources, including high-performance hardware and substantial memory.

These approach use labeled instances of the text to build a classifiers. Machine learning includes two approaches: Supervised learning and Unsupervised learning which in turn divided into several type[3].

3.2.1.1. Supervised learning

Supervised machine learning techniques training the dataset by using a labeled data set, each set composed of instances and each instance composed of probable entry and expected outcomes. Supervised machine learning is the most popular method because it's based on making the machine learn[18]. This technique have a several Certified approaches.

i. Decision tree classifier

Decision tree is one of the powerful methods in classification using for data mining purposes. Each node in the decision tree represents features to be classified[19]. It is a successive model, build a hierarchical tree structure and use training document for classifying based on queries of true and false[3]. The main pros is the Segmentation of data, but the Interpretability goes down when the number of splits increase. According to simple analysis on data forms, it implemented in many field. Entropy used in decision tree to measure a dataset's as in equation 1, where the value of entropy fall between 0 and 1[15].

$$Entropy(s) = \sum_{i=1}^{c} P_i \, log 2^{P_i} \qquad (1)$$

ii. Rule-based classifier

Rule-based classifiers are another type of classifier split data into set of rule which makes the class decision based on using different "if…else" rules. During the training phase generate rule in the form of "IF condition THEN conclusion", and the documents classify into categories. These rules are interpretable easier and the classifiers are utilize to create descriptive forms. The "if" condition is named the antecedent and expected class for every rule is named the consequent[3]. Rules can be created by utilize general-to-specific methods or specific-to-general methods. In general-to-specific methods, begin with rule with no antecedent and continue adding conditions until seeing enhancements in evaluation metrics, then, keep on eliminating the conditions of the rule covering pretty specific case[20].

iii. Support vector machine

SVMs were firstly evolves for classification and then extended to regression. SVM is a generalized linear supervised learning method used for classification. SVM minimizes the empirical classification mistake and increases the geometric margin. It is based on structural risk minimization[21] [22]. The output of the function is either positive or negative, so, SVM is considered a binary classifier, which discriminates data points of two categories and separates them with a hyperplane. SVM choose the hyperplane with the greatest margin to reach utmost separation among classes. To support nonlinear classification problems SVMs do the mapping from entry area to feature area [23].

iv. Neural network

This type is consists of neurons that linked to one another through weights that associate the neurons for easier transit of signals to revenue a single output[16]. The main advantages of neural network can execute tasks that a linear program unable to do it. It doesn't required to be re-programmed, it can be executed in any application. But the neural network require training to work [15].

v. Naive Bayes classifier

NB classifier are a set of classification algorithms depending on Bayes' Theorem. It's a group of algorithms where all of them sharing popular principle, i.e. each couple of features that classified is independent. This classification method is probabilistic method employ to predict the possibility for a given tuple is part of another class. Essentially the algorithm supposed that any feature of class is separate of other feature in the same class[22][24]. This technique commonly used for its simplicity through training and classification and it's based on Bayes theory, where the pre-processed data is the input to training set by classifier[16].

vi. Maximum entropy

It utilize probabilistic analysis and is appropriate for NLP. The MaxEnt algorithm utilize probability to locate the specific sentence part of which class, for that requires the uniform distribution and doesn't accept assumptions[22]. The MaxEnt is depending on the principle of Maximum Entropy, and selects which has the biggest entropy. The Max Entropy classifier can be utilize to solve a large textual classification problems such as detection the language, subject classification, and SA. Features can be added to MaxEnt without any overloading like phrases and bigram, also, Its utilize to assess probability distribution utilization logistic regression[16].

vii. Rough Set Theory (RST)

RST play a good role in SA via several languages in several applications such as political articles orientation[25] and political polarity detection[26], these researches using English and Arabic lexicon with RST. RST-based give a good results in Arabic/English sentiment analysis text. In addition to, RST used in English poem categorization which is part of SA in some time[27].

3.2.1.2. Unsupervised learning

The disadvantage of supervised methods is they need an enough amount of training data to getting accepted results. Upgrading such data is hard and costly process for that used the unsupervised learning methods to overcome the difficult of collection and establish labeled data[3][28].

3.2.2. Lexicon-Based approach

The lexicon-based approach compare the words in the text with the words in the maintained labeled list. The entry text is transform to tokens by tokenization process then every newly created token is identical in the dictionary and based on each token semantic orientation the overall sentiment of the text is computed, and the text is classified as positive, negative, or neutral[29]. The advantages of this approach:

1. Simplicity: Lexicon-based approaches are relatively straightforward to implement and understand. They rely on pre-defined word lists or dictionaries, making them easy to interpret and modify.
2. Adaptability: Lexicon-based approaches can be customized and tailored to specific domains or subject areas by creating or updating the lexicons accordingly. This adaptability makes them suitable for sentiment analysis, opinion mining, and other text analysis tasks.
3. Handling Out-of-Vocabulary Words: Unlike machine learning approaches, which may struggle with out-of-vocabulary words, lexicon-based approaches can handle such words by assigning sentiment based on the presence or absence of related words in the lexicon.

4. Interpretability: Lexicon-based approaches provide transparent results, as the sentiment analysis is directly linked to the predefined lexicons. This interpretability can be valuable in certain applications where understanding the reasoning behind sentiment analysis is crucial.

While the disadvantages are:

1. Limited Coverage: Lexicon-based approaches heavily depend on the quality and coverage of the lexicons or word lists. If the lexicon does not contain relevant words or lacks coverage in certain domains or languages, the accuracy and effectiveness of sentiment analysis may be compromised.
2. Contextual Ambiguity: may struggle with words or phrases that have multiple meanings or can be interpreted differently in different contexts. Assigning a sentiment to such ambiguous terms can lead to inaccurate results.
3. Lack of Learning and Adaptation: Unlike machine learning approaches, lexicon-based methods do not learn from data or adapt to changing patterns. They require manual updates or expansions of the lexicon to incorporate new words or sentiments.

Lexicon-based SA algorithm varies from other models in the manner that it clusters the sentiment values of positive and negative words in the text, when there is a matching, the score is incremented and the word is positive. The score is reduced and the word is tagged as negative [30]. 2 methods are utilize in Lexicon method the primary is a dictionary-based method and the other is the corpus-based method[3].

i. Dictionary approach

A dictionary is built manually with some basic words in a text editor. These words, collected inside a dictionary are specify with positive or negative values. The unknown and synonymous for words are found by utilize any online dictionary[30]. The dictionary utilize may be WordNet, secticNet, sentifull, SentiWordNet, or others. the antonyms, synonyms, and hierarchical used in opinion lexicons is to specifies word sentiments[3].

ii. Corpus approach

In the corpus-based approach, linguistic analysis and research are based on the study and analysis of large collections of text known as corpora. It is a structured and representative collection of texts, typically stored in electronic form, which serves as a linguistic resource for studying language patterns, usage, and properties. This approach rely on the statistical analysis of language data extracted from. This approach in sentiment analysis leverages the large scale analysis of text data to uncover sentiment patterns and develop models that can generalize sentiment predictions to new texts. Corpus-based Approach consist of two type Statistical approach and Semantic approach. Dictionary-based Approach are more efficient then Corpus Approach[31].

3.2.3. Hybrid approach

This technique combine machine learning and lexicon approaches. Researchers proved the combine of the two methods produces better perform of classification. The advantage of this method make the detection and scaling of sentiment at the conceptual stage, and the highly accuracy from the supervised learning algorithms[3][30][32].

4. Conclusion

It is notice that researchers upgrade their own frameworks for execution the classification algorithms in SA in various study cases, where most of frameworks include 4 stages involves; data collection, pre-processing, classification phase, and finally evaluation result. The SA techniques are arrange in three collections: Machine learning, Lexicon and hybrid approaches. The machine learning techniques are of two: supervised and unsupervised, which in turn is divided into a group of methods. In this paper an overview is presented about various classification techniques, their methods and the advantages, disadvantages of each type that used in SA on textual platform.

5. References

[1] V. Vaghela and B. M. Jadav, "Analysis of Various Sentiment Classification Techniques," *Int. J. Comput. Appl.*, vol. 140, no. 3, pp. 22–27, 2016.

[2] M. Ahmad, S. Aftab, I. Ali, and N. Hameed, "Hybrid tools and techniques for sentiment analysis: a review," *Int. J. Multidiscip. Sci. Eng.*, vol. 8, no. 4, pp. 28–33, 2017, [Online]. Available: www.ijmse.org

[3] *Sentiment Analysis Using Lexicon and Machine Learning-Based Approaches: A Survey*, vol. 34, no. May. Springer Singapore, 2018. doi: 10.1007/978-981-10-8198-9_46.

[4] S. Malviya, A. K. Tiwari, R. Srivastava, and V. Tiwari, "Machine Learning Techniques for Sentiment Analysis: A Review," *SAMRIDDHI A J. Phys. Sci. Eng. Technol.*, vol. 12, no. 02, pp. 72–78, 2020, [Online]. Available: www.ijmse.org

[5] N. C. Dang, M. N. Moreno-García, and F. De la Prieta, "Sentiment analysis based on deep learning: A comparative study," *Electron.*, vol. 9, no. 3, 2020, doi: 10.3390/electronics9030483.

[6] V. E. and G. Narayanan, "Sentiment Analysis for Hotel Reviews," *Proc. Comput. Linguist. Conf.*, vol. 231527, no. March, pp. 45–52, 2011.

[7] D. A. A and S. Pandey, "Sentiment Analysis Using Support Vector Machine Based on Feature Selection and Semantic Analysis," *Int. Res. J. Comput. Sci.*, vol. 8, no. 8, pp. 209–214, 2021, doi: 10.26562/irjcs.2021.v0808.009.

[8] V. Umarani, A. Julian, and J. Deepa, "Sentiment Analysis using various Machine Learning and Deep Learning Techniques," *J. Niger. Soc. Phys. Sci.*, vol. 3, no. 4, pp. 385–394, 2021, doi: 10.46481/jnsps.2021.308.

[9] A. Punitha, M. Tech, K. Kiruthigaipriya, B. Vanitha, and U. S. Harshanaa, "Sentiment Analysis for Depression based on Social Media Stream," vol. 8, no. 3, pp. 925–929, 2021.

[10] N. F. Al-Bakri, J. F. Yonan, A. T. Sadiq, and A. S. Abid, "Tourism companies assessment via social media using sentiment analysis," *Baghdad Sci. J.*, vol. 19, no. 2, pp. 422–429, 2022, doi: 10.21123/BSJ.2022.19.2.0422.

[11] S. Anis, S. Saad, and M. Aref, "Sentiment Analysis of Hotel Reviews Using Machine Learning Techniques," *Adv. Intell. Syst. Comput.*, vol. 1261 AISC, no. January 2020, pp. 227–234, 2021, doi: 10.1007/978-3-030-58669-0_21.

[12] A. K. Al-Mashhadany, A. T. Sadiq, S. M. Ali, and A. A. Ahmed, "Healthcare assessment for beauty centers using hybrid sentiment analysis," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 28, no. 2, pp. 890–897, 2022, doi: 10.11591/ijeecs.v28.i2.pp890-897.

[13] A. Katrekar, "An Introduction to Sentiment Analysis About me !!," *Adv. Soft Comput. Its Appl.*, 2019, [Online]. Available: http://link.springer.com/chapter/10.1007/978-3-642-45111-9_41

[14] M. D. Devika, C. Sunitha, and A. Ganesh, "Sentiment Analysis: A Comparative Study on Different Approaches," *Procedia Comput. Sci.*, vol. 87, pp. 44–49, 2016, doi: 10.1016/j.procs.2016.05.124.

[15] K. Mostafa, "A Survey of Sentiment Mining Techniques," *Int. J. Sci. Eng. Res.*, vol. 3, no. 3, pp. 1–6, 2015.

[16] M. Umar, "Sentiment Analysis Techniques and Application-Survey and Taxonomy," *Eng. Technol. J.*, vol. 04, no. 01, pp. 781–789, 2021, doi: 10.47191/etj/v6i1.01.

[17] Jason Brownlee, "Master Machine Learning Algorithms," 2019.

[18] E. Boiy and M. F. Moens, "A machine learning approach to sentiment analysis in multilingual web texts," *Inf. Retr. Boston.*, vol. 12, no. 5, pp. 526–558, 2009, doi: 10.1007/s10791-008-9070-z.

[19] B. Charbuty and A. Abdulazeez, "Classification Based on Decision Tree Algorithm for Machine Learning," *J. Appl. Sci. Technol. Trends*, vol. 2, no. 01, pp. 20–28, 2021, doi: 10.38094/jastt20165.

[20] W. Duch, N. Jankowski, K. Grąbczewski, and R. Adamczak, "Optimization and interpretation of rule-based classifiers," *Adv. Soft Comput.*, vol. 4, no. AISC, pp. 1–13, 2000, doi: 10.1007/978-3-7908-1846-8_1.

[21] D. K. Srivastava and L. Bhambhu, "Data classification using support vector machine," *J. Theor. Appl. Inf. Technol.*, vol. 12, no. 1, pp. 1–7, 2010.

[22] A. Sudhir Bale, S. Tiwari, I. Maknojia, S. Vura, and S. H. B, "Supervised Machine Learning Techniques for Sentiment Analysis and its Application in Image Processing and Remote Sensing," *Turkish Online J. Qual. Inq.*, vol. 12, no. 7, pp. 5759–5775, 2021, [Online]. Available: https://www.researchgate.net/publication/354048187

[23] G. Rozenberg, T. Back, and J. N. Kok, *Handbook of Natural Computing*, vol. 1–4, no. January 2012. 2012. doi:

10.1007/978-3-540-92910-9.

[24] J. Uma, “SENTIMENT ANALYSIS IN MACHINE LEARNING USING TWITTER DATA ANALYSIS IN PYTHON,” vol. 11, no. 12, pp. 3042–3053, 2020, doi: 10.34218/IJARET.11.12.2020.286.

[25] J. K. Alwan, A. J. Hussain, D. H. Abd, A. T. Sadiq, M. Khalaf, and P. Liatsis, “Political Arabic Articles Orientation Using Rough Set Theory With Sentiment Lexicon,” *IEEE Access*, vol. 9, pp. 24475–24484, 2021, doi: 10.1109/ACCESS.2021.3054919.

[26] D. H. Abd, A. R. Abbas, and A. T. Sadiq, “Twitter Sentiment Analysis and Events Prediction Using Lexicon Sentiment Analysis for Iraqi Vernacular,” *Bull. Electr. Eng. Informatics*, vol. 10, pp. 283–289, 2021.

[27] S. A. Alsaidi, A. T. Sadeq, and H. S. Abdullah, “English poems categorization using text mining and rough set theory,” *Bull. Electr. Eng. Informatics*, vol. 9, pp. 1701–1710, 2020.

[28] T. Zagibalov, “Unsupervised and Knowledge-poor Approaches to Sentiment Analysis,” no. May, p. 193, 2010, doi: 10.13140/RG.2.1.2488.7846.

[29] V. Bonta, N. Kumaresh, and N. Janardhan, “A Comprehensive Study on Lexicon Based Approaches for Sentiment Analysis,” *Asian J. Comput. Sci. Technol.*, vol. 8, no. S2, pp. 1–6, 2019, doi: 10.51983/ajcst-2019.8.s2.2037.

[30] P. Chauhan, “Sentiment Analysis: A Comparative Study of Supervised Machine Learning Algorithms Using Rapid miner,” *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. V, no. XI, pp. 80–89, 2017, doi: 10.22214/ijraset.2017.11011.

[31] M. Darwich, S. A. Mohd Noah, N. Omar, and N. A. Osman, “Corpus-Based Techniques for Sentiment Lexicon Generation: A Review,” *J. Digit. Inf. Manag.*, vol. 17, no. 5, p. 296, 2019, doi: 10.6025/jdim/2019/17/5/296-305.

[32] A. A. Tuama and A. T. Sadiq, “Twitter Sentiment Analysis and Events Prediction Using Lexicon Sentiment Analysis for Iraqi Vernacular,” *Solid State Technol.*, vol. 63, no. 6, pp. 10007–10015, 2020.

Article submitted 1 March 2023. Accepted at 10 Jun. Published at 30 Jun 2023.