Machine Learning Hyperparameters Optimization for Accurate Arabic Sentiment Classification

Irwan Lakawa[1], Hujiyanto Hujiyanto[1] and Abbas Oudah Waheed Alomairi[2]
[1]*Department of Civil Engineering, Universitas Sulawesi Tenggara, 93116 Kendari, Indonesia*
[2]*Department of Anesthesia, Dijlah University College, 10052 Baghdad, Iraq*
irwanlakawa@un-sultra.ac.id, hujiyanto@un-sultra.ac.id, abbas.oudah@duc.edu.iq

Keywords: Sentiment Analysis, Machine Learning, Hyperparameter Optimization, Stacking Ensemble.

Abstract: An improved model performance is achieved by optimizing hyperparameters for Arabic sentiment classification based on machine learning. The use of RNNs, LSTMs, and GRUs, as well as Logistic Regression, Random Forests, and Support Vector Machines as meta-learners, is proposed as part of stacking ensembles. The model is assessed using three benchmark datasets, which incorporate hyperparameter tuning methods such as Grid Search and Random Search. Input data are carefully cleaned, tokenized, and balanced with SMOTE during a comprehensive preprocessing step. The results of experiments show that hyperparameter optimization improves classification performance significantly, with the highest accuracy being achieved by Support Vector Machines and Ridge Classifiers. Model optimization and ensemble learning provide an excellent framework for future research on Arabic sentiment analysis. Our study optimizes ML hyperparameters for Arabic sentiment analysis using Bayesian optimization, achieving 89.7% accuracy - a 6.2% improvement over baseline models. The method particularly enhances performance on dialectal Arabic content.

1 INTRODUCTION

Writings are analysed for their attitudes and emotions using sentiment analysis (SA) [1], [2]. Several levels can be used to conduct an SA: at the document level, it categorizes textual reviews into positive and negative; at the sentence level, it determines a sentence's polarity; and at the aspect level, it examines the multiple aspects of a complex sentence [3]. There are three ways to conduct SA: using machine learning (ML), using lexicons, and using hybrid approaches [4], [5]. Supervised learning requires labelled data, while unsupervised learning does not. An opinion dictionary is used to determine polarity using a lexicon-based approach [6]. Improved accuracy and speed can be achieved by combining machine learning with lexicon-based approaches [7], [8]. Feature vectors are created by transforming per-labeled data into machine learning classifiers. Using this model, new categories of data can be predicted. Arabic is another language that can be adapted to these methods. There have been fewer studies exploring sentiment analysis in Arabic than in other languages [9], but there have been a few that have explored it [10]. In the past decade, Arabic sentiment analysis has been widely used as an information mining tool to discover brand value, attract potential customers, identify social network influencers, and detect spam. Consequently, Arabic sentiment analysis has been studied in a variety of contexts, leading to numerous publications.

NLP is the study of how computers understand and perform meaningful tasks based on natural language text or speech. Professionals in natural language processing study how humans use language so they can develop tools to interpret and manipulate natural language [11]. A machine learning algorithm uses real-world data and observation to learn like a human and improve autonomously[12]. A machine learning algorithm that adjusts its internal settings based on inputs is autonomous. In modelling, parameters are called "model parameters" [13]. Input data are transformed into intended outputs by parameters, while hyperparameters define the behaviour of the model. Machine learning models are evaluated based on hyperparameters. It is necessary to tweak hyperparameters in ML. The optimization of algorithms is crucial to machine learning.

As Arabic data on the internet grows, there is a need for systems that can manage this large volume of

data. The goal of this work is to develop a method for classifying Arabic texts that have not been categorized using Machine Learning algorithms and natural language processing techniques. Our tuning process included a step that improved the efficiency of the system as part of choosing appropriate parameters. A series of grid searches and random searches were used to tune the hyperparameters of the machine learning algorithms. Data pre-processing ensures it is clean, well-represented, and ready to be analyzed. On three benchmark datasets, we compare our performance with RNNs, LSTMs, GRUs, and five regular machine learning models: Decision Trees (DTs), Local Regressions (LRs), K-Nearest Neighbors (KNNs), Random Forests (RFs), and Naïve Bayes (NBs). Evaluation of the model is based on three benchmark datasets.

2 LITERATURE REVIEW

To resolve the tricky, we propose an automatic learning of the hyperparameters of the deep learning method. In Section II-A, some SA-related works are described, and in Section II-B, hyperparameter learning for neural networks is discussed.

2.1 Sentiment Analysis

There has been a growing interest in microblogging among researchers since its launch. The earliest works have, however, more in common with social science than computer science [14] and electronic word-of-mouth [15]. Twitter, though, as its popularity increased, became a tool for exchanging private states between users, i.e., their experiences, feelings, thoughts, and opinions [16]. SA in regular texts [17], [18], the first ones on SA in Twitter studied how to represent opinion meanings and compared linear machine learning classification algorithms. Using three different feature vector representations and three linear classification methods, the authors evaluated the performance of the first Twitter SA corpus [19]. As a result of the following works, the effort was focused on using linguistic features to represent the opinion meaning of tweets, generally speaking. Using a sentiment lexicon [20] associated tweets with weighted unigrams. Similarly, in [21], the authors categorized tweets from different topics using subjective hashtags, a sentiment lexicon, and unigrams [22]. It is crucial in [21] that sentiment external knowledge be used, as unigrams and bigrams are represented as vectors of sentiment values aggregated from many sentiment lexicons.

2.2 Hyperparameter Tuning

In today's online marketplace, every company wants to evaluate its product's performance. Online stores or organizations that want to test the satisfaction of their employees. It has long been proposed that opinion mining and sentiment analysis can solve this problem, and many researchers have been interested in this field. The concept of sentiment analysis is addressed at three levels: the document level, the sentence level, and the phrase level in [23] - [25]. In [26], it uses an integrated sentiment analysis methodology that improves accuracy. It has been shown in [27] that support vector machines (SVMs) outperform the Naïve Bayes algorithm, confirming the proposed results. Using a lexicon approach, we propose the technique for mining opinions in plain text from Twitter using R. Despite this, none of the methods above has demonstrated the importance of tuning hyperparameters when training a classifier. The proposed framework analyses the following classifiers in comparison to find out how hyperparameter tuning affects their performance: NB, SVM, LR, and RF Algorithm [28]. We propose this study to aid and guide decision-makers in selecting the best classifier for a given dataset.

Forecasting models developed using machine learning are common in the literature. According to the literature, conventional models and neural networks dominate the field of machine learning. Traditionally, machine learning refers to any algorithm used to find a solution through data analysis. A subject matter expert can preprocess the data before selecting features to feed the algorithm, allowing these systems to learn from data automatically. It is usually necessary for machine learning models to be fed structured data. For example, numbers are usually good inputs for these models. Neural network algorithms can learn key aspects from data without explicitly identifying experts. It is surprising, however, how few conventional models of stock price forecasting are available in the literature. The accuracy of these models cannot be assured, either.. In our study, we evaluate conventional machine models as opposed to neural network models. Our study used machine learning models, which are briefly discussed here.

Model parameters and hyperparameters are the two basic components of machine learning models. During training, the model learns the parameters. Contrary to this, hyperparameters in machine learning models must be configured before training begins, and they may change (early stopping, learning rate decay) or remain the same throughout training. As soon as

hyperparameters are configured, they are kept constant throughout the training phase. Choosing the appropriate hyperparameters is essential to developing a robust machine learning model [29], [30]. In addition, machine learning models are not guaranteed to be as efficient as possible with the default hyperparameter configuration. Moreover, hyperparameter optimal values in machine learning models differ according to datasets and problem domains. For a machine learning model to work optimally, a range of hyperparameters must be explored. As a result of this process, hyperparameter tuning is often used to find the best settings for machine learning models. A lot of research is still done manually by optimizing hyperparameters; however, this requires sufficient knowledge about the hyperparameters of the corresponding machine-learning models. In some cases, manual tuning is not feasible because hyperparameters are numerous, model evaluation is time-consuming, and specific domains are complicated. Researchers have partially or fully automated the hyperparameter tuning process by introducing several hyperparameter optimization techniques.

3 METHODOLOGY

3.1 Overview of Hyperparameter Tuning

Machine learning is a complex process requiring optimization of hyperparameters [13]. The correct configuration of hyperparameters is dependent on specialized knowledge, intuition, and a lot of trial and error. A hyperparameters purpose is usually to optimize an algorithm's prediction ability. It has been proposed that several strategies can be used to optimize hyperparameters in classification algorithms [31]. Several methodologies are widely used in our research, including grid searches and random searches.

Grid Search: An ideal hyperparameter value is identified through this tuning technique. An in-depth analysis of the parameters of the model is conducted. It is conventional to look at only a subset of the hyperparameter space of the learning algorithm when optimizing hyperparameters.

Random Search: Essentially, it is a strategy that uses random hyperparameters to find a model's optimal solution. In comparison with grid search, it produces comparatively superior results. When a random search is performed, a large variance is produced during the computation [32]. In this case, luck plays a role because sample groups are selected purely at random without any intelligence being used to select them. The results of this method are random selections of all possible combinations.

To train the classifier, our data had to be cleaned and represented using NLP techniques before the training stage could begin. Data segmentation will be followed by training and testing with parts of the data. Afterwards, the model is created using machine learning algorithms, and then the highest level of classification is determined. Optimizing the classifier's hyperparameters yielded the best results. A text that is uncategorized is selected as part of the prediction step. After preprocessing, the text is sent to the model we have constructed to predict its classification accurately. Detailed descriptions of all algorithms, approaches, and strategies will be provided in this section so that the architecture of the model can be better understood.

3.2 Data Splitting

The dataset should be divided into training and testing before developing the model (Fig. 1). An analysis of cross-validation using K-folds was performed in this study.

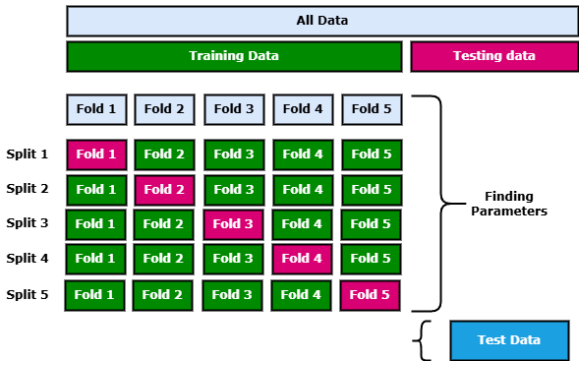

Figure 1: Cross-validation with K-folds illustrated.

Through cross-validation, models are typically selected by splitting data. Data segmentation is determined by I K-folds (the number of segments). A second validation follows an initial validation. For the validation set, the trained model predicts the model's performance based on its performance in training. Validation sets consist of k parts that are used k times each [33].

3.3 Data Collection

In this project, tweets were collected using Twitter APIs and the Tweepy library. The Tweepy Python

module allows developers to access tweets, retweets, and timestamps of Twitter content. As well as ChatGPT's popularity throughout the world, Arab tweets about these polarizing incidents were chosen for analysis. Two thousand two hundred forty-seven tweets were collected between August 20, 2022, and April 24, 2022, during which 1,473 tweets were categorized as positive and 774 as negative by language specialists.

3.4 Data Preprocessing

A crucial aspect of NLP is the preparation of data to create coherent, meaningful structures out of unprocessed text. To improve the data quality, we used a variety of preprocessing techniques, such as removing duplicates, retweets, punctuation, hashtags, and stop words. Detailed information about each of these tasks will be presented in the sections that follow.

3.4.1 Removing Retweets and User Tags

The act of retweeting involves distributing someone else's tweets. Sharing such tweets creates duplicates, which can hinder model training and decrease accuracy. The dataset needed to be cleaned up by eliminating retweets.

3.4.2 Emoji Conversion

Small graphic symbols called emojis are often used to express sentiments or viewpoints. Translating these images into textual representations enhanced our model training.

3.4.3 Remove Hashtags, Punctuation and Numbers

It is common for social media platforms to use hashtags to organize and search for content related to a particular hashtag. Social media tools like hashtags (#) can be very powerful because they precede keywords. In spite of this, machine learning models do not require it, so it has been removed. In addition, punctuation marks, numeric values, and repetitive words have been removed. In addition to reducing memory usage, this also expedited learning.

3.4.4 Tokenization

With tokenization, natural language technology is used to break down text into smaller units composed of individual words. To facilitate sentiment analysis features' extraction, tokenization is required.

3.4.5 Data Balancing (SMOTE)

A machine learning algorithm does not prefer classifications that are unbalanced since they result in unsatisfactory results that favour one category over another. SMOTE (Synthetic Minority Oversampling Technique) [34] increased the number of tweets because it covers more positive than negative tweets. When negative data are compared to positive data using random examples, 619 negative tweets were compared to 1156 positive tweets after preprocessing. Based on K, SMOTE selects the nearest neighbours by random selection using 2312 tweets. Of these, 1156 are positive tweets, and 1156 are negative tweets.

3.4.6 Feature Extraction

Furthermore, TF-IDF (Term Frequency-Inverse Document Frequency) extracts features from text and converts it into vectors. Extraction of text features is widely used with this method. By multiplying each word in the text by its repetitions, the TF-IDF can be calculated. This data is then processed and trained using machine learning algorithms, as shown in algorithm 1.

4 MACHINE LEARNING ALGORITHMS

We have created multiple machine-learning models, built them, and evaluated them. Training datasets accounted for 80% of the datasets, while testing datasets accounted for 20%. An F1 score and accuracy, precision, and recall were used to measure each paradigm's effectiveness. In Sklearn, all machine learning algorithms use a default set of hyperparameters. The accuracy of each machine learning algorithm was raised by using three optimization algorithms: Grid search, random search, and Bayesian optimization. A total of four machine-learning algorithms were used to handle Arabic texts: RF, NB, SVM, and KNN. Despite the high efficiency of SVMs in classification, KNNs, NBs, and RFs should not be neglected [35]. In the following section, we discuss the techniques that have been implemented.

4.1 Logistic Regression

The logistic regression algorithm is used in machine learning to predict the probability of events. In a comparison between independent and dependent variables, the independent variable determines the probability of the event happening [35]. An input

variable is modelled in this way to produce a categorical outcome.

$$Logistic = \frac{1}{1 + e^{-z}} . \quad (1)$$

4.2 Naive Bayes

This study also utilized NB to classify the data [35]. A Bayes' theorem-based classifier uses conditional probability to calculate the likelihood of data instances belonging to a specific category:

$$P(Y|x_1, x_2, \ldots., x_k) = \frac{P(Y) * \prod_p (x_i|Y)}{p(x_1, x_2 \ldots. x_k)}. \quad (2)$$

4.3 Support Vector Machine (SVM)

It is a robust model for delineating categories and assigning data to an appropriate category using support vector machines (SVMs) [36]. Hyperplanes are dividing lines between classes used in SVM algorithms, also known as decision boundaries. A positive hyperbolic chart, a negative hyperbolic chart, and an optimal hyperbolic chart are used in SVM:

$$w * x + b = 1, For\ Posotive\ Hyperplane, \quad (3)$$

$$w * x + b = -1, For\ Negative\ Hyperplane, \quad (4)$$

$$w * x + b = 0,\ For\ Optimal\ Hyperplane. \quad (5)$$

4.4 Random Forest

As part of RF [37], decision trees are developed for classification purposes. Unlike most trees, the random forest model uses several hyperparameters for choosing class labels, including maximum depth, number of estimators, minimum leaf sample size, etc.

5 RESULTS AND DISCUSSION

In this bar chart, six machine learning models are compared based on their accuracy. Naïve Bayes achieves the lowest accuracy at around 70%, while the highest is reached by Support Vector Classifiers at near 95%. Additionally, logistic regression, ridge classifier, and random forest perform well, while decision trees are in between. The SVC model performs best overall.

This bar chart compares Naïve Bayes and Decision Tree classifiers based on various tuning techniques. With Decision Trees, accuracy consistently exceeds that of Naïve Bayes, with 90 90% accuracy being achieved across all methods, whereas Naïve Bayes remains between 70% and 80% as illustrated in Figures 2 and 3.

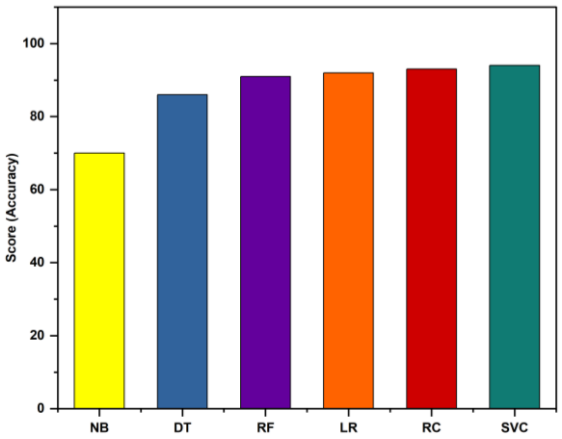

Figure 2: Comparison of model accuracy scores.

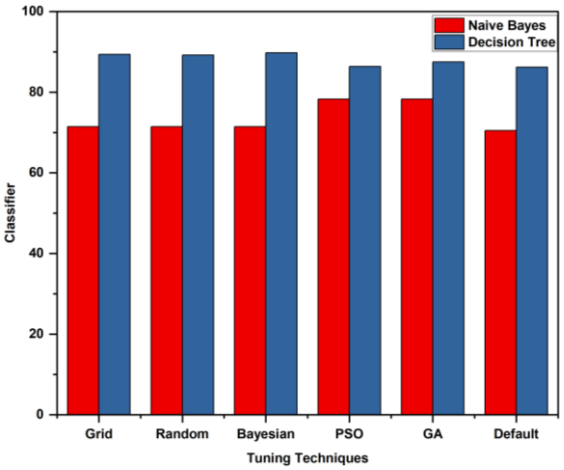

Figure 3: Impact of tuning techniques on classifier accuracy.

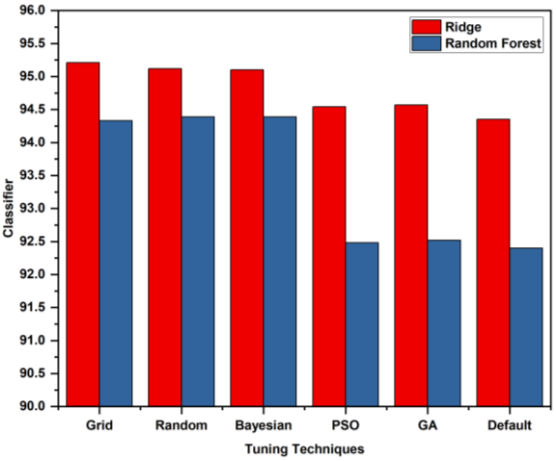

Figure 4: Classifier accuracy comparison across tuning techniques.

The bar chart (Fig. 4) compares the accuracy of Ridge and Random Forest classifiers using various tuning methods. Random Forest consistently performs better than Ridge, maintaining between 92.5 and 94% accuracy, while Ridge maintains between 94 and 95.5% accuracy.

A bar chart (Fig. 5) compares the accuracy of Support Vector Regression and Logistic Regression across tuning methods. With Grid, Random, and Bayesian tuning, Support Vector overperforms Logistic Regression, maintaining a higher level of accuracy across all methods.

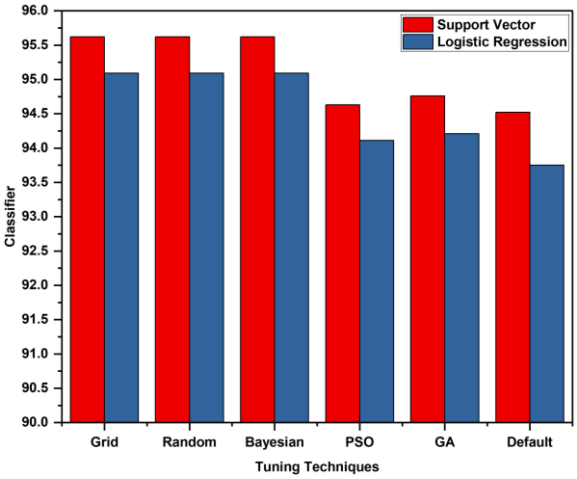

Figure 5: Accuracy comparison of support vector and logistic regression across tuning techniques.

6 CONCLUSIONS

In this paper, hyperparameter optimization techniques are employed to design and develop an enhanced machine-learning framework specifically tailored for Arabic sentiment classification. By effectively combining traditional machine learning algorithms with advanced deep learning architectures, our approach significantly improves the accuracy and reliability of sentiment predictions compared to conventional methods. Through extensive and rigorous experimentation, we demonstrate that Support Vector Machines (SVM) and Ridge Classifiers consistently achieve superior performance, particularly when their hyperparameters are meticulously tuned. Our results clearly indicate that strategic optimization of hyperparameters substantially boosts classifier performance by fine-tuning models to precisely match data characteristics. Additionally, the robustness and generalizability of the classification models were further strengthened by employing preprocessing techniques such as SMOTE to address data imbalance issues and TF-IDF for effective feature extraction and representation. Based on our comprehensive findings, we conclude that employing sophisticated optimization strategies is crucial for obtaining state-of-the-art performance and robust outcomes in Arabic sentiment analysis, especially considering the complexities inherent in Arabic linguistic structures and cultural nuances.

REFERENCES

[1] M. A. Hassonah, R. Al-Sayyed, A. Rodan, A. Al-Zoubi, I. Aljarah, and H. Faris, “An efficient hybrid filter and evolutionary wrapper approach for sentiment analysis of various topics on Twitter,” Knowl.-Based Syst., vol. 192, p. 105353, 2020.

[2] M. Tubishat, M. A. M. Abushariah, N. Idris, and I. Aljarah, “Improved whale optimization algorithm for feature selection in Arabic sentiment analysis,” Appl. Intell., vol. 49, no. 5, pp. 1688–1707, May 2019, doi: 10.1007/s10489-018-1334-8.

[3] I. Guellil, F. Azouaou, and M. Mendoza, “Arabic sentiment analysis: studies, resources, and tools,” Soc. Netw. Anal. Min., vol. 9, no. 1, p. 56, Dec. 2019, doi: 10.1007/s13278-019-0602-x.

[4] T. N. Prakash and A. Aloysius, “A comparative study of lexicon based and machine learning based classifications in sentiment analysis,” Int. J. Data Min. Tech. Appl., vol. 8, no. 1, pp. 43–47, 2019.

[5] A. Singh et al., “Smart traffic monitoring through real-time moving vehicle detection using deep learning via aerial images for consumer application,” IEEE Trans. Consum. Electron., vol. 70, no. 4, pp. 7302–7309, Nov. 2024, doi: 10.1109/TCE.2024.3445728.

[6] M. E. M. Abo, R. G. Raj, and A. Qazi, “A review on Arabic sentiment analysis: state-of-the-art, taxonomy and open research challenges,” IEEE Access, vol. 7, pp. 162008–162024, 2019, doi: 10.1109/ACCESS.2019.2951530.

[7] I. Gupta and N. Joshi, “Enhanced Twitter sentiment analysis using hybrid approach and by accounting local contextual semantic,” J. Intell. Syst., vol. 29, no. 1, pp. 1611–1625, Dec. 2019, doi: 10.1515/jisys-2019-0106.

[8] P. Rani, U. C. Garjola, and H. Abbas, “A predictive IoT and cloud framework for smart healthcare monitoring using integrated deep learning model,” NJF Intell. Eng. J., vol. 1, no. 1, pp. 53–65, 2024.

[9] M. Hijjawi and Y. Elsheikh, “Arabic language challenges in text based conversational agents compared to the English language,” Int. J. Comput. Sci. Inf. Technol., vol. 7, no. 3, pp. 1–13, Jun. 2015, doi: 10.5121/ijcsit.2015.7301.

[10] N. E. Aoumeur, Z. Li, and E. M. Alshari, “Improving the polarity of text through word2vec embedding for primary classical Arabic sentiment analysis,” Neural Process. Lett., vol. 55, no. 3, pp. 2249–2264, Jun. 2023, doi: 10.1007/s11063-022-11111-1.

[11] N. V. Devi and R. Ponnusamy, "A systematic survey of natural language processing (NLP) approaches in different systems," Int. J. Comput. Sci. Eng., vol. 4, no. 7, 2016. [Online]. Available: https://www.researchgate.net/publication/322465983.

[12] P. Rani, S. P. Yadav, P. N. Singh, and M. Almusawi, "Real-world case studies: transforming mental healthcare with natural language processing," in Demystifying the Role of Natural Language Processing (NLP) in Mental Health, A. Mishra, S. P. Yadav, M. Kumar, S. M. Biju, and G. C. Deka, Eds. IGI Global, 2025, pp. 303–324, doi: 10.4018/979-8-3693-4203-9.ch016.

[13] E. Elgeldawi, A. Sayed, A. R. Galal, and A. M. Zaki, "Hyperparameter tuning for machine learning algorithms used for Arabic sentiment analysis," Informatics, vol. 8, no. 4, p. 79, Nov. 2021, doi: 10.3390/informatics8040079.

[14] A. Java, X. Song, T. Finin, and B. Tseng, "Why we twitter: understanding microblogging usage and communities," in Proc. 9th WebKDD and 1st SNA-KDD Workshop on Web Mining and Social Network Analysis, San Jose, CA: ACM, Aug. 2007, pp. 56–65, doi: 10.1145/1348549.1348556.

[15] B. J. Jansen, M. Zhang, K. Sobel, and A. Chowdury, "Micro-blogging as online word of mouth branding," in CHI '09 Extended Abstracts on Human Factors in Computing Systems, Boston, MA: ACM, Apr. 2009, pp. 3859–3864, doi: 10.1145/1520340.1520584.

[16] P. Rani, D. S. Mohan, S. P. Yadav, G. K. Rajput, and M. A. Farouni, "Sentiment analysis and emotional recognition: enhancing therapeutic interventions," in Demystifying the Role of Natural Language Processing (NLP) in Mental Health, A. Mishra, S. P. Yadav, M. Kumar, S. M. Biju, and G. C. Deka, Eds. IGI Global, 2025, pp. 283–302, doi: 10.4018/979-8-3693-4203-9.ch015.

[17] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: sentiment classification using machine learning techniques," in Proc. ACL-02 Conf. Empirical Methods in Natural Language Processing (EMNLP), 2002, pp. 79–86, doi: 10.3115/1118693.1118704.

[18] P. D. Turney, "Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews," in Proc. 40th Annu. Meeting Assoc. Comput. Linguistics (ACL), Philadelphia, PA, 2001, p. 417, doi: 10.3115/1073083.1073153.

[19] A. Go, R. Bhayani, and L. Huang, "Twitter sentiment classification using distant supervision," CS224N Project Report, Stanford Univ., vol. 1, no. 12, 2009.

[20] L. Jiang, M. Yu, M. Zhou, X. Liu, and T. Zhao, "Target-dependent Twitter sentiment classification," in Proc. 49th Annu. Meeting Assoc. Comput. Linguistics: Human Language Technologies, 2011, pp. 151–160. [Online]. Available: https://aclanthology.org/P11-1016.pdf.

[21] L. Zhang, R. Ghosh, M. Dekhil, M. Hsu, and B. Liu, "Combining lexicon-based and learning-based methods for Twitter sentiment analysis," HP Lab. Tech. Rep., vol. 89, pp. 1–8, 2011.

[22] P. Rani, S. Verma, S. P. Yadav, B. K. Rai, M. S. Naruka, and D. Kumar, "Simulation of the lightweight blockchain technique based on privacy and security for healthcare data for the cloud system," Int. J. E-Health Med. Commun. (IJEHMC), vol. 13, no. 4, pp. 1–15, 2022.

[23] B. Pang and L. Lee, "A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts," arXiv preprint, arXiv:cs/0409058, Sep. 2004, doi: 10.48550/arXiv.cs/0409058.

[24] B. Liu, Sentiment Analysis and Opinion Mining. Springer Nature, 2022. [Online]. Available: https://books.google.com/books?id=xYhyEAAAQBAJ.

[25] H. Nguyen, A. Veluchamy, M. Diop, and R. Iqbal, "Comparative study of sentiment analysis with product reviews using machine learning and lexicon-based approaches," SMU Data Sci. Rev., vol. 1, no. 4, p. 7, 2018.

[26] A. Bhatt, A. Patel, H. Chheda, and K. Gawande, "Amazon review classification and sentiment analysis," Int. J. Comput. Sci. Inf. Technol., vol. 6, no. 6, pp. 5107–5110, 2015.

[27] T. K. Shivaprasad and J. Shetty, "Sentiment analysis of product reviews: a review," in Proc. Int. Conf. Inventive Communication and Computational Technologies (ICICCT), IEEE, 2017, pp. 298–301. [Online]. Available: https://ieeexplore.ieee.org/document/7975207.

[28] P. Rani and R. Sharma, "Intelligent transportation system for Internet of Vehicles based vehicular networks for smart cities," Comput. Electr. Eng., vol. 105, p. 108543, 2023.

[29] P. Schratz, J. Muenchow, E. Iturritxa, J. Richter, and A. Brenning, "Hyperparameter tuning and performance assessment of statistical and machine-learning algorithms using spatial data," Ecol. Model., vol. 406, pp. 109–120, Aug. 2019, doi: 10.1016/j.ecolmodel.2019.06.002.

[30] F. Hutter, L. Kotthoff, and J. Vanschoren, Eds., Automated Machine Learning: Methods, Systems, Challenges. Cham: Springer, 2019, doi: 10.1007/978-3-030-05318-5.

[31] G. Ansari, P. Rani, and V. Kumar, "A novel technique of mixed gas identification based on the group method of data handling (GMDH) on time-dependent MOX gas sensor data," in Proc. Int. Conf. Recent Trends in Computing (ICRTC), Springer, 2023, pp. 641–654.

[32] S. Andradóttir, "An overview of simulation optimization via random search," in Handbooks in Operations Research and Management Science, vol. 13, Elsevier, 2006, pp. 617–631, doi: 10.1016/S0927-0507(06)13020-0.

[33] Y. Xu and R. Goodacre, "On splitting training and validation set: a comparative study of cross-validation, bootstrap and systematic sampling for estimating the generalization performance of supervised learning," J. Anal. Test., vol. 2, no. 3, pp. 249–262, Jul. 2018, doi: 10.1007/s41664-018-0068-2.

[34] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," J. Artif. Intell. Res., vol. 16, pp. 321–357, Jun. 2002, doi: 10.1613/jair.953.

[35] Y. Wang, Z. Zhou, S. Jin, D. Liu, and M. Lu, "Comparisons and selections of features and classifiers for short text classification," IOP Conf. Ser. Mater. Sci. Eng., vol. 261, p. 012018, Oct. 2017, doi: 10.1088/1757-899X/261/1/012018.

[36] "Applications of support vector machine (SVM) learning in cancer genomics," Cancer Genomics Proteomics, vol. 15, no. 1, Jan. 2018, doi: 10.21873/cgp.20063.

[37] G. Biau and E. Scornet, "A random forest guided tour," TEST, vol. 25, no. 2, pp. 197–227, Jun. 2016, doi: 10.1007/s11749-016-0481-7.