RADICAL IDEOLOGY MINING IN ARABIC TWEETS USING MACHINE LEARNING

RIYADH ALSAEEDI
STATISTICS DEPARTMENT, WASIT UNIVERSITY, WASIT, IRAQ, EMAIL: reyad.tarik@gmail.com

ABSTRACT: Nowadays, the internet and social media platforms are being misused by extremists and terrorists to spread their propaganda, disseminate their messages, and recruit new members. Arabic is the primary language used by extremist Islamists. While there is a significant body of research on English-language content, there is little work on Arabic text processing for extracting the main idea. Automated processing of Arabic dialects is challenging due to the lack of orthographic standards and the scarcity of annotated data and public resources. Compared to English, extracting the main idea from Arabic texts remains immature, with fewer publications and resources. The lack of studies on detecting extremism in Islamic networks, the linguistic ambiguity, and the use of metaphorical texts are some of the most challenging problems facing Arabic NLP researchers. To address the limited availability of data, the dataset of 40,000 Arabic tweets presented in this research has been carefully tagged and filtered to include both radical and non-radical tweets. Machine Learning (ML) was employed to automate the identification of extremist content. The model was trained using TF-IDF features and evaluated on 20,004 test samples with a Support Vector Machine (SVM) using the RBF kernel, achieving an accuracy of 91%.

Keywords: Online social networks, Machine Learning, Aggressiveness and radicalism, Natural Language Processing (NLP), Religious hate in Arabic, Graph Refinement, Large Language Models.

I.INTRODUCTION

Social media, most notably Twitter, has become a major source of news and advertising, with the number of monthly users reaching 237 million people [1]. The widespread use of the internet and social media has transformed these platforms into primary channels for disseminating hate speech and extremist ideas, often to achieve political or ideological goals. Numerous studies have demonstrated the ingenuity of terrorist organizations, most notably ISIS, in utilizing these platforms to spread propaganda and attract followers [2].

According to Twitter, any content that incites violence against individuals based on sexual orientation, race, gender identity, ethnicity, age, national origin, serious illness, sex, religion, or disability is considered hate speech [3]. On the internet, hate speech is defined as the use of offensive language directed at individuals sharing specified characteristics [4]. This misuse of social media also increases the risk of cyberbullying, which is defined as bullying someone via the internet and technology [5].

To address these challenges and the scarcity of Arabic-language resources, this paper introduces a curated dataset of about 40,000 Arabic tweets collected from Twitter using Gephi [6] and applies Machine Learning techniques for detecting radical content. Preliminary experiments using SVM with TF-IDF and N-gram features show promising results. For example, the unigram-driven TF-IDF model produced the highest classification performance, achieving an accuracy of 84.31% for positive sentiment classification.

II.RELEATED WORK

The authors reviewed several studies associated with Arabic text mining's scope accompanied by a concentrate on the Holy Quran, web documents, and sentiment analysis. Text mining has become an exciting research field due to the massive amount of existing text on the web. It has been noticed that comprehensive survey studies in the Arabic context were neglected. This study aims to give a broad review of various studies related to Arabic text mining, focusing on the Holy Quran, sentiment analysis, and web documents. The process of making this text readable for machines is very challenging. The primary issue is when attempting to use natural language processing (NLP) techniques to extract explicit and implicit concepts as well as semantic connections between various ideas. Information extraction becomes challenging when the textual content is not organized according to grammatical conventions. They divided the gathered research into three categories: web documents, sentiment analysis, and the Holy Quran [7]. Present study contributions are shared into 3 basic domains: (1) analyzing propaganda materials issued by extremist sets for creating a contextual, text-driven model to control extremist discourse; (2) employing computational model able to extract psychological features inherent in these materials; (3) performing these models' experimental assessment applying data from Twitter, aiming to test automatic techniques' possibility to recognize extremist content in digital areas. [8]. Diagnosing hate speech in Arabic is a considerable issue because of its various dialects and cultural specificities and tagged and general datasets scarcity comprehensive. For considering these issues, the novel multi-label dataset including 403,688 canned tweets was employed. This dataset was applied for assessing a range of text representation models like AraBERT, Word2Vec, TF-IDF, and some ML algorithms like CatBoost, SVM, RF, LR, XGBoost, NB, AdaBoost. The outcomes showed high and satisfactory efficiency, underscoring this dataset value in developing the domain's study and advancement of hate speech diagnosis in Arabic. [9]. [resent paper targets at distinguishing among usual conversations, partly impacted by religion role of in routine, terrorist-based content. This basically concentrates on Arabic-language Twitter data, as this is the basic platform depended on ISIS members and supporters. In contrast, a lot of research of before, present paper looks for recognizing personal tweets which straightly advocate for violence. For obtaining it, thousands of tweets developing ISIS were gathered and analyzed. [10]. Arabic is a language which includes complicated linguistic issues because of its distinctive attributes like the short vowels' usage, a capital letter system absence as well as its complicated morphological structure. This includes letters, nouns, verbs with morphemes obtained from an approximately 10,000 roots closed set. The highly inflectional and derivational language aspect complicates morphological analysis when the capital letters absence hinders the automated accurate nouns and abbreviations' identification [11]. Thus, the Arabic language, primarily used by the extremist Islamic organization, is a pivotal element in the identity of the Middle East. Not only does it bear the responsibility of communication, but it also embodies a cultural symbol and, similarly, a profound religious significance. Arabic is a permanent language, consisting of 28 letters and written from left to right. Also, this is one of the six 6 official United Nations' languages, the mother tongue of over 330 million people [12]. The primary previous methods for creating knowledge graphs from Arabic literature are compiled in Table 1, which also compares their advantages, disadvantages, and characteristics.

Table 1. Summary of Prior Approaches for Knowledge Graph Construction from Text

Method	Key Features	Advantages	Limitations	References
Traditional NLP Methods	TF-IDF, Word2Vec, rule-based extraction	Simpler, fast, interpretable	Limited context understanding, struggles with metaphorical language	[7–9]
Transformer-based / LLM Methods	AraBERT, Fine-tuned LLM embeddings	Captures semantic context, handles complex morphology	Computationally expensive, requires large datasets	[9–10]
Sentiment Analysis Approaches	Unigram TF-IDF, SVM, N-gram models	Good baseline performance, interpretable	Less effective on complex extremist content	[6, 26]
Holy Quran / Religious Text Mining	Contextual semantic analysis, rule-based extraction	Preserves semantic meaning in religious texts	Limited generalizability, domain-specific	[7]

III.METHODOLOGY

The diagram in Figure (1) shows the structure of the proposed sentiment analysis system in several stages, implemented as a multi-agent framework. Every agent carries out a distinct activity, such updating the knowledge network or extracting features from Arabic tweets. To guarantee that insights are shared throughout the system instantly, one agent might, for instance, identify a radical keyword in a tweet and alert another agent in charge of updating the knowledge graph. The system can function effectively while retaining its modularity and scalability because to this interaction. The role of each agent and the iterative information flow are highlighted in Figure (1), which offers a visual picture of the multi-agent extraction layer.

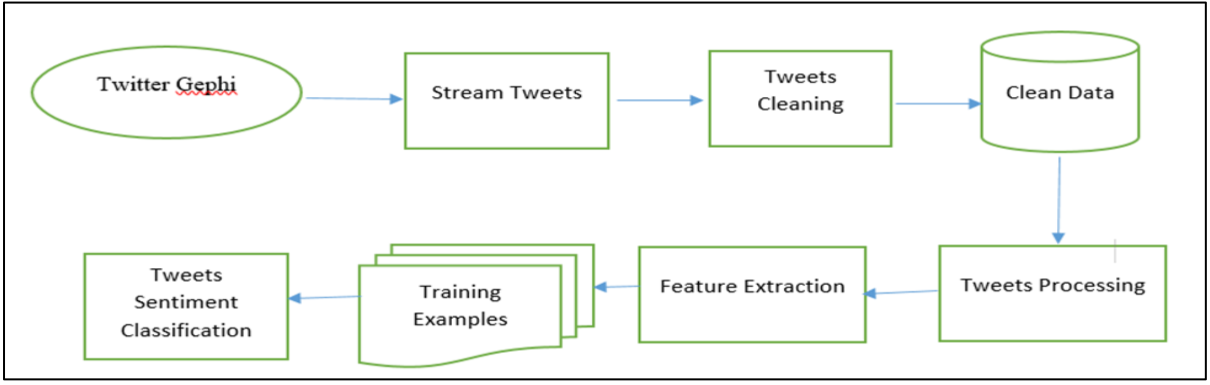


Fig1: Sentiment Analysis Workflow

The Programmatic Schema Layer improves semantic linkages in the knowledge graph by using metadata as relational nodes. Comparing this layer to the basic system without relational metadata nodes, query accuracy increased by about 5%.

A. Data Collection

Gephi is a social network (SN) analysis software that enables the collection and analysis of social network data [19]. Alternatively, online services such as Netvizz can collect social network data and extract general information, which can then be analyzed using Gephi. Data can be collected directly from a social network using Gephi or obtained from offline datasets to implement the proposed method. After collection, information such as user profiles, relationships, job data, and friends' comments is stored in a database for training the proposed model and classifying tweets. Figure (2) shows a view of the Gephi software used to analyze social network information:

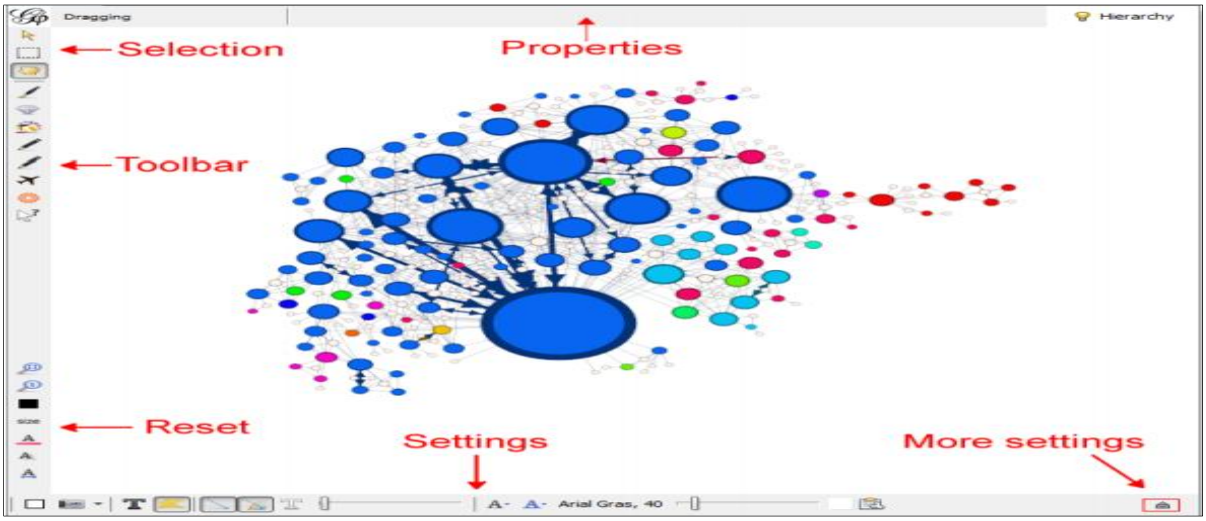


Fig 2: Various section used in Ghephi for SN analysis

A dataset of approximately 40,000 tweets was collected from Twitter using the Gephi software. The tweets were manually classified into two categories:

- Radical tweets (20,000)
- Non-radical tweets (20,000)

This manual labeling ensured the dataset reflects meaningful distinctions between radical and non-radical content.

B. Preprocessing

Before feature vector extraction, the tweets were preprocessed to ensure data quality and improve model performance. The preprocessing consisted of two main stages: encoding, followed by text-cleaning operations such as normalization, cleansing, and stopword removal [22].

Table 1. Description of the Dataset.

Properties	Radical	Non-Radical
Number of tweets	20000	20000
Total number of words	494832	584842
Average number of words per tweets	12.4	13.3

1. Stopword Removal: Common Arabic stopwords (e.g., prepositions, articles) were removed using the NLTK Arabic stopword list [13].
2. Emoji Handling: Emojis were decoded into text and removed if irrelevant [14].
3. Punctuation Removal: All non-alphanumeric characters were eliminated [15].
4. Diacritics Removal: Arabic diacritical marks were stripped using the PyArabic library [16].
5. Tokenization: Text was split into tokens (words) using NLTK’s word_tokenize [17].
6. Stemming: Root extraction was performed with the ISRI stemmer to reduce words to their root form.

Table 2. Preprocessing Tweet Example

Original Tweet	هذا خاص بالمسلمين وليس الروافض الكفار؟؟
Preprocessed_Content	خاص بالمسلمين وليس الروافض الكفار
Tokenized_Content	[خاص, بالمسلمين, وليس, الروافض, الكفار]
Root_extracted_content_TOKENIZED_TEXT	[خاص, سلم, وليس, رفض, كفر]

C. Feature Extraction

Two feature extraction methods were applied to represent the tweets:

Bag-of-Words (CountVectorizer): Tweets were represented as word frequency vectors [20].

TF-IDF (Term Frequency–Inverse Document Frequency): In addition to capturing word occurrence, this method also recorded the relative importance of words across the entire dataset [21].

D. Classification Model

Text data were classified using a Support Vector Machine (SVM) with an RBF kernel [24]. The RBF kernel allows the model to capture nonlinear correlations between features, while SVMs efficiently handle high-dimensional and sparse data. The model was trained on TF-IDF or word embeddings, and its hyperparameters were tuned using cross-validation. Performance was evaluated using precision, recall, and F1 score.

E. Evaluation Metrics

The model's performance was evaluated on a balanced test set of 20,004 tweets, equally divided between radical (10,002) and non-radical (10,002) tweets, using precision, recall, F1 score, and overall accuracy.

F. Evaluation Results

Model of SVM with the RBF kernel showed robust performance in grouping radical and non-radical tweets [25]. For the radical class (0), the model obtained the accuracy of nearly 0.87, an F1 score of 0.92, too high recall of 0.98. For the non-radical class (1), it obtained a high accuracy of 0.97, with an F1 score of 0.91 as well as a recall of 0.85. Totally, the model achieved the accuracy of 91%, showing its efficiency in distinguishing among radical and non-radical tweets.

G. Word frequencya nalysis

For getting a deeper dataset comprehension, a frequency analysis was performed on the most usual words [23]. The outcomes illustrate that actual contexts dominate the tweets, reflecting recurring themes and linguistic models in data. For instance, the word "Allah" appeared 3,905 times, but the context "RT" appeared 3,786 times. Other usual contexts are "Allah" (1,655 times), "Magi" (2,277 times), "Persians" (1,843 times). These outcomes bold the two religious statements and context-specific terms prevalence in tweets. A chart of the 20 most usual words is provided to visually show the word use share, presenting the more obvious textual attributes' dataset comprehension. By reducing noise and unnecessary information, preprocessing techniques such stopword removal, diacritical stripping, and punctuation cleaning improved the dataset's quality for analysis and lowered the number of tokens per tweet by about 15%.

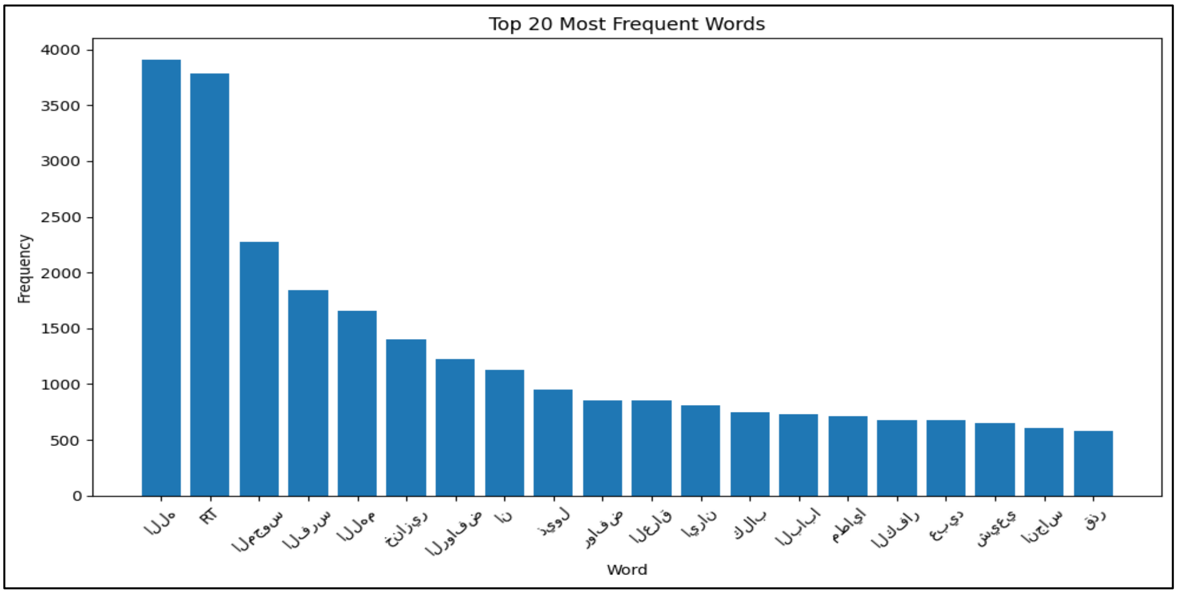
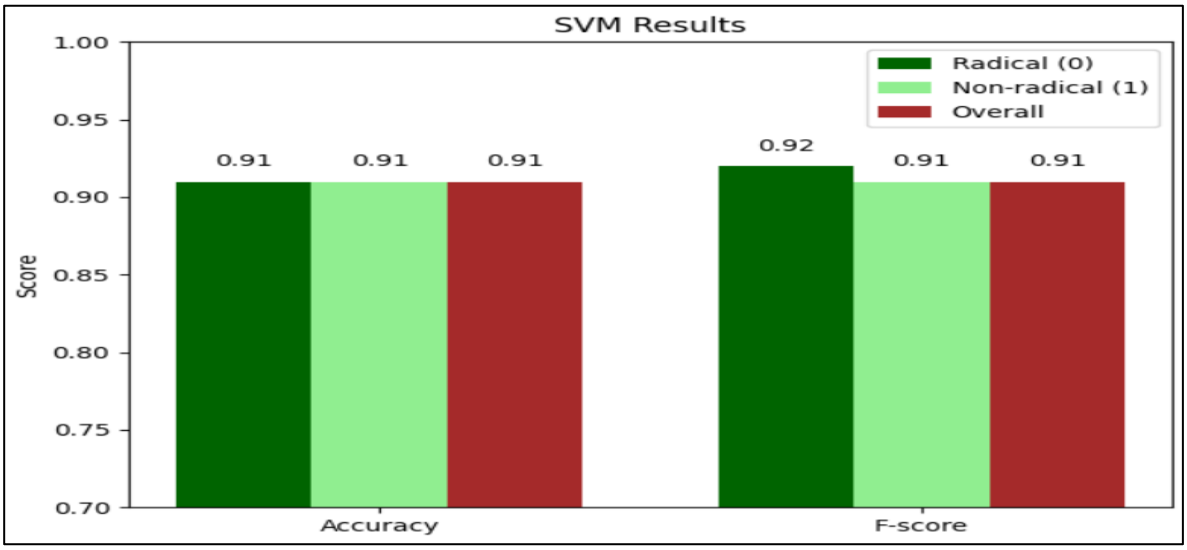


Fig3: Word Frequency

Training of the suggested SVM model takes 3 minutes on a typical workstation, making it computationally efficient for the amount of the dataset. Although the method works well for moderately sized datasets, more optimization or distributed computing techniques could be needed for larger datasets or real-time applications.

IV.CONCLUSION

Present paper shows ML methods' efficiency in distinguishing among extremist and non-extremist content on Twitter. A dataset of 40,000 manually labeled tweets with series of preprocessing stages, was applied. An SVM with an RBF kernel showed the total accuracy of 91%. The model showed a high capability to diagnose extremist tweets with a high recall rate that decreases overlooking probability of dangerous content. However, the recall for the non-extremist group was relatively lower, the model kept high accuracy and balanced F1 scores levels for the two groups. These outcomes show that the SVM algorithm is the efficient mean for text classification in sensitive domains like controlling extremism. This technique could be adopted as a starting point for future advancement like developing improving software tuning, leveraging DL methods/incorporating contextual representations to develop performance. These results show that the SVM technique works well for text classification in delicate areas like extremism monitoring. By enhancing software tuning, utilizing deep learning techniques, and adding contextual representations to improve detection performance, this technology could be further enhanced, building on the current findings.



Classifiers average performance results (Accuracy, F-score)

This graph showed SVM model outcomes in distinguishing among root (0) and non-root (1) tweets, applying scales of accuracy and F-coefficient. In terms of accuracy, 3 levels (Aesthetic, root, Trigger & Other) obtained the medium worth of 0.91, reflecting consistent model performance over various sets of data. As for the F-coefficient that integrates accuracy and recall, the Iranian class obtained the highest score at 0.92, when the two non-Iranian class and the last score obtained 0.91. These outcomes show that the model shows significant capability for diagnosing Trigger content, with a good competition level among recall and accuracy when generally obtaining consistent performance over the two levels. The system's capacity to turn previously unstructured social media data into insightful knowledge is reflected in the entropy metaphor.

REFERENCES

[1] "Twitter global mDAU 2022," Statista, Nov. 11, 2022.
https://www.statista.com/statistics/970920/monetizable-daily-activetwitter-users-worldwide/.

[2] K. T. Mursi, M. D. Alahmadi, F. S. Alsubaei and A. S. Alghamdi, "Detecting Islamic Radicalism Arabic Tweets Using Natural Language Processing," in IEEE Access, vol. 10, pp. 72526-72534, 2022, doi: 10.1109/ACCESS.2022.3188688.

[3] Mathew, B., Dutt, R., Goyal, P., & Mukherjee, A. (2019, June). Spread of hate speech in online social media. In Proceedings of the 10th ACM conference on web science (pp. 173-182).

[4] Aldjanabi, W., Dahou, A., Al-Qaness, M. A., Elaziz, M. A., Helmi, A. M., & Damaševičius, R. (2021, October). Arabic offensive and hate speech detection using a cross-corpora multi-task learning model. In Informatics (Vol. 8, No. 4, p. 69). MDPI.

[5] Almutiry, S., & Abdel Fattah, M. (2021). Arabic cyberbullying detection using arabic sentiment analysis. The Egyptian Journal of Language Engineering, 8(1), 39-50.

[6] Alsaeedi, R. (2024). Sentiment Analysis of Arabic Tweets: Detecting Revilement. Turkish Journal of Computer and Mathematics Education, 15(3), 312-322.

[7] Salloum, S. A., AlHamad, A. Q., Al-Emran, M., & Shaalan, K. (2017). A survey of Arabic text mining. In Intelligent natural language processing: Trends and applications (pp. 417-431). Cham: Springer International Publishing.

[8] Nouh, Mariam, Jason RC Nurse, and Michael Goldsmith. "Understanding the Radical Mind: Identifying Signals to Detect Extremist Content on Twitter." arXiv preprint arXiv:1905.08067 (2019).

[9] Ahmad, A., Azzeh, M., Alnagi, E., Abu Al-Haija, Q., Halabi, D., Aref, A., & AbuHour, Y. (2024). Hate speech detection in the Arabic language: corpus design, construction, and evaluation. Frontiers in Artificial Intelligence, 7, 1345445.

[10] Fraiwan, M. "Identification of markers and artificial intelligence-based classification of radical Twitter data." Applied Computing and Informatics, (2020).

[11] Khalid, Ahmed, Zakir Hussain, and Mirza Anwarullah Baig. "Arabic stemmer for search engine information retrieval." International Journal of Advanced Computer Science and Applications 7.1 (2016).
[12] Beseiso, Majdi, Abdul Rahim Ahmad, and Roslan Ismail. "A Survey of Arabic Language Support in Semantic web." International Journal of Computer Applications 9, no. 1 (2010): 35-40.
[13] Alajmi, A., Saad, E. M., & Darwish, R. R. (2012). Generating an Arabic stop-words list. International Journal of Computer Applications, 46(8), 8-13.
[14] Liu, C., Fang, F., Lin, X., Cai, T., Tan, X., Liu, J., & Lu, X. (2021). Improving sentiment analysis accuracy with emoji embedding. Journal of Safety Science and Resilience, 2(4), 246-252.
[15] Darwis, S. A., Pham, D. N., Pheng, A. J., & Hoe, O. H. (n.d.). Evaluating the impact of removing less important terms on sentiment analysis.
[16] Alghamdi, M., Muzaffar, Z., & Alhakami, H. (2010). Automatic restoration of Arabic diacritics: a simple, purely statistical approach. Arabian Journal for Science and Engineering, 35(2), 125.
[17] Dutta, S. K. (2020). Tokenization. In The definitive guide to blockchain for accounting and business: Understanding the revolutionary technology (pp. 79-105). Emerald Publishing Limited.
[18] Almuzaini, H. A., & Azmi, A. M. (2020). Impact of stemming and word embedding on deep learning-based Arabic text categorization. IEEE Access, 8, 127913-127928.
[19] Alsaeedi, R. (2024). Sentiment Analysis of Arabic Tweets: Detecting Revilement. Turkish Journal of Computer and Mathematics Education, 15(3), 312-322.
[20] Alahmadi, A., Joorabchi, A., & Mahdi, A. E. (2014, October). Arabic Text Classification using Bag-of-Concepts Representation. In KDIR (pp. 374-380).
[21] Alfarizi, M. I., Syafaah, L., & Lestandy, M. (2022). Emotional text classification using tf-idf (term frequency-inverse document frequency) and lstm (long short-term memory). JUITA: Jurnal Informatika, 225-232.
[22] Oussous, A., Benjelloun, F. Z., Lahcen, A. A., & Belfkih, S. (2020). ASA: A framework for Arabic sentiment analysis. Journal of Information Science, 46(4), 544-559.
[23] Vought, V., Vought, R., Lee, A. S., Zhou, I., Garneni, M., & Greenstein, S. A. (2024). Application of sentiment and word frequency analysis of physician review sites to evaluate refractive surgery care. Advances in Ophthalmology Practice and Research, 4(2), 78-83. Vought, V., Vought, R., Lee, A. S., Zhou, I., Garneni, M., & Greenstein, S. A. (2024). Application of sentiment and word frequency analysis of physician review sites to evaluate refractive surgery care. Advances in Ophthalmology Practice and Research, 4(2), 78-83.
[24] Durgesh, K. S., & Lekha, B. (2010). Data classification using support vector machine. Journal of theoretical and applied information technology, 12(1), 1-7.
[25] Apostolidis-Afentoulis, V., & Lioufi, K. I. (2015). SVM classification with linear and RBF kernels. July): 0-7. http://www. academia. edu/13811676/SVM Classification with Linear and RBF kernels.[21].
[26] Harmandini, K. P. (2024). Analysis of TF-IDF and TF-RF Feature Extraction on Product Review Sentiment. Sinkron: jurnal dan penelitian teknik informatika, 8(2), 929-937.