INTERNATIONAL JOURNAL OF TECHNO-ENGINEERING
IJTE

ISSN: 2057-5688

A Novel Machine Learning-based Framework for Detecting Religious Arabic Hatred Speech in Social Networks

S. Sailaja [1], G. Bhargavi [2]
[1] Associate professor, Associate Professor [2]
Department of Computer Science Engineering
RISE Krishna Sai Prakasam Group of Institutions

Abstract—Social media platforms generate a huge amount of data every day. However, liberty of speech through these networks could easily help in spreading hatred. Hate speech is a severe concern endangering the cohesion and structure of civil societies. With the increase in hate and sarcasm among the people who contact others over the internet in this era, there is a dire need for utilizing artificial intelligence (AI) technology innovation that would face this problem. The rampant spread of hate can dangerously break society and severely damage marginalized people or groups. Thus, the identification of hate speech is essential and becoming more challenging, where the recognition of hate speech on time is crucial in stopping its dissemination. The capacity of the Arabic morphology and the scarcity of resources for the Arabic language makes the task of distinguishing hate speech even more demanding. For fast identification of Arabic hate speech in social network comments, this work presents a comprehensive framework with eight machine learning (ML) and deep learning (DL) algorithms, namely Gradient Boosting (GB), K-Nearest Neighbor (K-NN), Logistic Regression (LR), Naive Bayes (NB), Passive Aggressive Classifier (PAC), Support Vector Machine (SVM), Ara-BERT, and BERT-AJGT are implemented. Two representation techniques have been used in the proposed framework in order to extract features: a bag of words followed by BERT-based context text representations. Based on the result and discussion part, context text representation techniques with Ara-BERT and BERT-AJGT outperform all other ML models and related work with accuracy equal to 79% for both models.

Keywords—Machine learning; Arabic language; hatred detection; social network; classification algorithm

I. INTRODUCTION

Low-resource languages, e.g., Arabic, Hindi, and Urdu, do not have a considerable amount of data for training and building conversational Artificial Intelligence (AI) systems. The Arabic language is the authorized language for 22 Arabic countries with roughly more than 422 million aboriginal speakers [1]. Additionally, it is a religious language spoken by more than 1.5 billion Muslims. There are three main types of it: i) Classical Arabic which is the

ISSN: 2057-5688

language of the Holy Quran, ii) Modern Standard Arabic (MSA) which is used by academia, and iii) Dialectical Arabic which differs between regions since it is used for daily life networking [2]. Thus, the Arabic language, with a large number of speakers worldwide, is challenging task when we work with AI systems. Moreover, the Arabic language is identified as the 4 th in the usage on the internet [3]. However, the sophistication of the Arabic language makes the automatic identification of Arabic hate speech a complex task. Dialectal Arabic doesn't have formal grammar or spelling regulations. Moreover, spelled words can have different importance based on various dialects, which augments the vagueness of the language [4]. Sociable applications, e.g., Facebook, YouTube, and Twitter, are generating an extensive quantity of data which is considered a valuable goldmine for researchers. Social mediagenerated data helps in recognizing unlawful behavior, restricting potential hurt, and maintaining residents safe [2]. Some users utilize the wild adoption of online social networks to spread radical and biased statements that diffuse hate speech. Sentiment analysis (SA), i.e., opinion mining, analyses individuals' thoughts, attitudes, emotions, and opinions towards an entity, e.g., person, object, or service. The SA is implemented at various levels of granularity [5]: i) document-level: each text fragment is considered as a component with an opinion towards a single object. It aims to categorize a review as positive, negative, or neutral, ii) sentence-level: aims to extract opinions for a smaller text which is more challenging than SA for a document, and iii) aspect-level: determines the major features of a belief while focusing on aspect extracting and feeling categorization of aspects. The approaches of semantic analysis could be supervised, unsupervised, or hybrid. For unsupervised methods, numerous sentiment phrases are required to reveal the semantic orientation of texts. Thus, lexiconbased approaches are used. However, supervised techniques rely mostly on utilizing data mining tools to build a learning algorithm on a collection of tagged information. A prime application of SA is hate speech detection through online social network [6]. Hate speech is any class of inappropriate language, e.g., insults, slurs, threats, encouraging violence, and impolite language, that targets individuals or groups based on typical attributes such as nationality, religion, ideology, disability, social class, or gender. Hate speech includes racism, misogyny, religious discrimination, and abusive speech. Racism implies hate speech that attacks people based on their skin color, race, origin, class, or nationality [7]. Misogyny is the hate speech that targets females, i.e., women or girls [8]. Religious bias is hatred vocabulary towards somebody based on their beliefs, faiths, practices, or even the deficiency of religious faiths The abusive speech represents disrespectful, rude, or criticizing speech to hurt or deliver harmful sentiments. Hate speech (HS) detection is a branch of offensive language detection. There is an increasing studies for

ISSN: 2057-5688

abusive/HS detection for English language. However, it is still very limited for Arabic dialects due to the scarcity of the publicly obtainable resources required for abusive/HS detection in Arabic social media texts. The authors of [9] declared that the harmful online content on social media can be grouped into various categories including: Vicious, Vulgar, Offensive, Violent, Adult content, Terrorism and Spiritual hate speech. This work targets religious hate speech (RHS) that could be insulting, abusive, or hateful. RHS aims to instigate hate, intolerance, or roughness toward people because of their religious faiths. The recent immense usage of social networks mandates applying different text processing on such cyberspace. The remarkable amount of generated data requires applying new monitoring tasks such as cyberbullying recognition [10], hate speech detection [6], irony identification [11], and discovery of offensive language [12]. Accordingly, battling hate speech mandates generating and elucidating a considerable amount of data for automatic hatred speech identification by building artificial intelligence-based models, i.e., ML and DL [13]. Lately, detecting abusive and hateful speech has gained increasing attraction from investigators in NLP and computational social sciences societies. Thus, detecting abusive speech and hate speech is essential for online safety. lately, various studies indicated that the existence of hate speech may be related to hate crimes [8]. Therefore, this work aims to enhance the detection of offensive language and hate speech on Arabic text. Detecting religious hate speech in any language, including Arabic, has different challenges, including : 1) the gigantic volume of the data generated over social networks makes it difficult to locate typical patterns and trends in the data, 2) noise may exist in the data, e.g., inaccurate grammar, misspelled phrases, Internet slang, abbreviations, lengthening of words, and multi-lingual scripts, 3) the comments being written in poorly text, and including paralinguistic signs, e.g., emoticons, and hashtags. Moreover, hate detection is a contextdependent task, and it is still missing a consense of what is forming hate speech due to the different cultures, customs and traditions, and 4) since the social networks prevent posting illegal content, users post information that looks authentic and simple but quietly causes a hate speech. Thus, building a tool for the automatic detection of hate speech would be complex [6]. Social platforms, e.g., Twitter and Facebook began battling online hate speech by explaining procedures that limit the use of violent and dehumanizing languages [14]. Moreover, various Arabic countries, where their users of social media sites are adding Arabic content, modified their laws to combat cybercrimes including hate speech. For example, Jordan added a new cybercrime laws [15] that defines hate speech as any action, writing, or speech planned to cause and raise ethical conflict or call for violence and provocation to fighting between the diverse segments of the nation. Regarding

ISSN: 2057-5688

the Arabic language, there is a clear shortage in the conducted research for hate speech on online social networks. Thus, artificial intelligence, data mining, and machine learning techniques could be utilized to efficiently perform more research and experiments on hate speech detection which constitutes a fertile resource for investigation. This work aims to design a prototype for the automatic identification of abusive and hate speech using various ML and DL techniques with a standard data set.

II. PRELIMINARIES

A. Natural Language Processing (NLP)

Natural Language Processing is a major component of Artificial Intelligence (AI). It enables robots to analyse and comprehend human language, enabling them to carry out repetitive activities without human intervention. Machines can analyse and comprehend human language through a process known as NLP. NLP-based approaches process a considerable amount of data to obtain useful knowledge. For that different data mining and machine learning approaches are used. Thus, text pre-processing should be applied in order to prepare text for further processing such as representation features engineering that are required to extract features and pass it to ML approaches. For example, pre-processing could include text tokenization, and stop-word removal.

B. Machine Learning (ML) Algorithms

ML is used in various applications, e.g., healthcare [16], hardware design [17], quality control [18], and NLP, where this work targets NLP application. Information is an organised collection of discrete pieces of data, and it conceals the whole spectrum of representational patterns. The machine's primary objective is to extract patterns that reflect a certain event. If the machine is able to recognise these patterns, then machine learning has taken place. It demonstrate that by adding fresh data or information, where the computer can make accurate predictions. The authors of [19] have mentioned that the advancements in machine learning especially deep learning enable us to design algorithms that use realworld information to make decisions that seem subjective. As shown in Section IV-B, there are different methods to prepare text for further processing. Text tokenization, which is also called text segmentation or lexical analysis, groups the text into tokens/words separated by space. Stop-words such as articles (e.g., a, an, the), conjunctions (e.g., and, but, if), and prepositions (e.g., in, at, on) [20], do not represent a specific meaning. Thus, they should be eliminated. Features in ML are essentially numerical attributes. However, the data may not contain numerical attributes, such as in sentiment analysis. Thus, various types of features (e.g., word, character, so on) are converted into numerical features where such operation is called representation and choosing from them which make ML working properly is

ISSN: 2057-5688

called feature engineering (feature selection and feature extraction).

C. Hate Speech

Recently, the broad usability of smartphones and the high availability of internet access increased the number of users on social media. Moreover, the rapid growth of social media has made it practically unattainable to manually monitor and inspect the massive amount of messages published online every day. Also, social media witnessed a substantial increase in hate and abusive speech, which is a severe problem worldwide that threatens the solidarity of civil communities. Therefore, automatic detection for hate speech, utilizing various classification techniques, is required to filter such harmful content. Twitter is one of the most importing social media platform which is ubiquitous, informal, and unstructured at the same time. Tweets usually have abbreviations, acronyms, spelling errors, and nonideal punctuation so designing a model to handle this will be an interesting topic for future work.

D. Transfer Learning (TL)

ML still has some constraints for specific real-world domains. For example, the requirement of having a tremendous amount of training data which have a distribution similar to the testing data could be difficult to satisfy [21]. Thus, semisupervised learning could be utilized due to the shortage of labeled data. However, for a small amount of unlabeled data, the build model would be defective. Therefore, transfer learning is a promising procedure for such systems. Transfer learning (TL) is a branch of machine learning (ML) which aims to improve the performance of target learners on specific fields by transferring the knowledge possessed in separate but connected source domains [21]. Thus, constructing target learners will have a reduced dependency on a large number of targetdomain data. In ML models, knowledge is not retained or accumulated, where learning is performed without considering past learned knowledge in other tasks. However, in transfer learning, the learning process can be faster, more accurate, and require less training data. TL can be classified into: 1) homogeneous where the disciplines are of the identical feature space, 2) heterogeneous where the disciplines have diverse feature spaces.

E. Data Oversampling and Underselling (Re-Sampling)

With the tremendous increase in the size of the generated data in various applications, there is a lack of equality in the labeled data. However, various ML techniques assume equal distribution for the target classes which is not always a realistic assumption. Such class imbalance problems will have a good accuracy while other evaluation metrics

ISSN: 2057-5688

including precision, recall, F1-score, and ROC (Receiver Operating Characteristics) score, will not have enough scores. As shown in Fig. 1, Resampling including under-sampling or oversampling could be used to resolve the problem of an imbalanced data set. Undersampling reduces the amount of the majority target samples. On the other hand, oversampling raises the quantity of minority class instances by yielding new instances or reproducing some instances [22].

III. RELATED WORK

Various researches have been conducted to detect hate speech as a wide notion with different types in the English language. Many proposed works performed hate speech Fig. 1. Undersampling vs Oversampling [22]. detection as a binary classification problem and considered a broad concept such as detecting bullying and derogatory language. In [23], the authors presented an original technique to detect hatred speech in English tweets. For that, they utilized three models, i.e., logistic regression (LR), XGBoost classifier (XGB), and support vector machine (SVM). The obtained performance showed competitive results compared to standard stacking, base classifiers, and majority voting techniques. The authors of [24] determined and discussed challenges encountered by online automatic techniques for hate speech detection in text. The limited availability of the data, sensitivity in language, and the exact definition of what forms of hate speech are well-known challenges. They proposed a SVM technique with high performance while the decisions are easier to interpret than neural methods. However, the used datasets did not include Arabic text. In [25], the authors used different machine learning algorithms for the automatic identification of hate speech in tweets written in the Indonesian language. Their results showed that the Multinomial Naive Bayes algorithm has the most promising results with a value of 71.2% and 93.2% for accuracy and recall, respectively. The authors of [2] researched the capability of deep learning based on Convolutional Neural Networks (CNN), CNN-long short-term memory networks (CNN-LSTM), and bidirectional LSTM (BiLSTM-CNN) to automatically detect hateful content posted on social media. For that, they used the ArHS dataset with 9833 tweets, which is believed to be the largest Arabic dataset with hate speech content. The authors of [14] aimed to identify Cyber hate speech within the Arabic content of Twitter where they used various NLP and ML techniques. In [26], the authors used Twitter to construct an Arabic text detection hate speech model. They use this knowledge to analyze a dataset of 11 thousand tweets. They apply the Term Frequency — Inverse Document Frequency (TF-IDF) words representation to the SVM model. Finally, they presented four deep learning models that can notice and classify Arabic hate speech on Twitter into several types. In [27], the authors were the

ISSN: 2057-5688

first who addressed the problem of recognizing speech encouraging religious hatred in the Arabic Twitter. Thus, they were able to detect messages that use provocative sectarian speech to promote hatred and violence against people based on their religious beliefs. They found that a simple Recurrent Neural Network (RNN) architecture with Gated Recurrent Units (GRU) can adequately detect religious hate speech. The used data set is available online at [28]. The authors of [29] presented the foremost publicly-available Levantine Hate Speech and Abusive (L HSAB) Twitter dataset. It is intended to be a benchmark dataset for automatic detection of online Levantine harmful contents. The dataset, which is a available at [30], includes 5,846 tweets that could be of Normal class, Abusive, or Hate speech. Considerable work has been investigated for hatred speech detection in the English language. However, rare work has targeted the detection of hate speech in the Arabic language. The majority of the Arabic research targeted web pages and search engines, while a few targeted comments on social networks. In this work, we target the Arabia language and use the data set of [28]. Thus, our constructed models would be mainly compared with [27].

IV. PROPOSED METHODOLOGY

The proposed architecture for Arabic hate speech detection is showing in Fig. 2. It includes the subsequent major steps: collection of labelled text document/tweet, text preprocessing, text representation and feature extraction, building of classification models (learning), and Relearning (testing) and classification process.

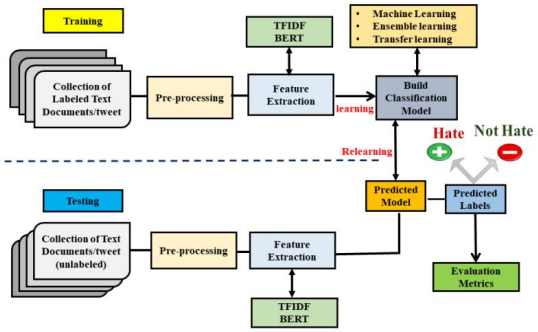

Fig. 2. The Proposed Architecture for Hate Speech Detection Model.

A. Data set: Collection of Labelled Text Document/Tweet

In this work we used the data set which was collected by [27] and it is available online at [28]. The data set contains 6164 Arabic tweets and concentrates on the four most typical sacred religions in the Middle East, which are Islam (93.0%), Christianity (3.7%), Judaism (1.6%), and Atheism (0.6%). Originally, the training data set contains 5,569 examples, while the testing data set contains 567 document. The data works for binary classification with two hateful and non hateful classes represented by 1 and 0 respectively. Since data resampling is be utilized to settle the issue of an imbalanced data set, we performed re-sampling technique. According, the model built with data oversampling is called Classifier-Over while the model built with

ISSN: 2057-5688

data under sampling is called Classifier-Under, where Classifier could be any of the six models we used.

B. Data Pre-Processing

Text pre-processing includes various techniques that prepare text for further processing. Pre-processing aims to remove the unwanted words from the text, e.g., punctuation, slang, and stop words. Usually, we have to deal with various preprocessing techniques and a combination of them, including: 1) Tokenization: Tokenization is the activity of splitting text into terms, phrases, symbols or additional important elements, called tokens [31]. The obtained elements can be single items (1-gram) or a series of n words (n-gram). Items can be phonemes, syllables, letters, words or even sentences. 2) Stopwords Removal: Our work targets Arabic text. Thus, for pre-processing stage, we first remove the nonArabic text. Every non-Arabic character is replaced with a whitespace character. Moreover, we remove stopwords, which appear frequently in the text and are not important for text classification, e.g., ©Ó , ð @ , ú ¯ , úÎ« , á« , áºË . A list of the most frequently used Arabic stop words is available at [20]. Approximately, 20%–30% of the total words in a record are stopwords, that is, terms that can be removed as they are redundant without any semantic value [32]. The traditional approach for extracting stopwords includes a pre-filled list, containing all words that are semantically irrelevant to a specific language. This technique is a static. On the other hand, the stopwords are recognized online and not specified previously for the dynamic technique. The features are specified based on their importance. Similar to the removal of stopwords, this work eliminates the punctuation and digits from the Arabic text.

V. EXPERIMENTAL ANALYSIS

A 2 × 2 matrix, which is called a confusion matrix, is created to visually illustrate the performance of a binary supervised learning problem. Table I shows the confusion matrix for Arabic hate speech detection. It includes four classes, which are true positive, true negative, false positive, and false negative. In this work, True Positive (TP) indicates that the comment is actually hate speech and correctly classified as hate speech. True Negative (TN) indicates that the comment is non-hate speech and correctly classified as non-hate speech. False Positive (FP) means that the comment is actually nonhate speech but incorrectly classified as hate speech, and False Negative (FN) describes the comment that is actually hate speech but incorrectly classified as non-hate speech. For any classification model, we aim to maximize the value of TP and TN and minimize the value of FP and FN.

ISSN: 2057-5688

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall/Sensitivity = \frac{TP}{TP + FN}$$

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

we will explain the result of various machine learning models which we have used in this study. We have designed different models to detect and classify religious Arabic hate speech based on various methods, e.g., data partitioning, re-sampling and transfer learning. For that, we have divided our data (train/test) for three scenarios. The data is partitioned into (70/30), (80/20), and (90/10). For each partitioning, we use the original data sets in addition to the oversampling and under-sampling techniques. The bestobtained classification performance in partitioning scenarios was for (80/20), where a detailed explanation is given in Section VI-A. Then, Section VI-B explains the classification performance for 70/30 data partitioning while Section VI-C is dedicated to 90/10 data partitioning. The proposed models were evaluated on a testing data set related to religious hate speech in Arabic text [28]. In order to enhance the performance of the classifiers that we build for six ML algorithms, we have extended this implementation to include transfer learning methods. The transfer learning models called Ara-BERT and AJGT-BERT. The comparison of the performances of all models have been done in terms of accuracy, recall, precision, and F1 score using Arabic hate speech data set.

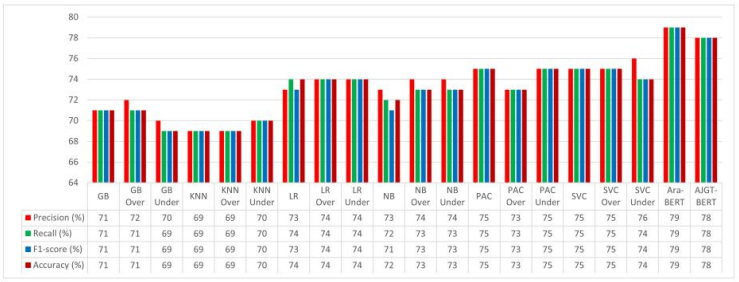

Fig. 3. Various Classification Metrics for 80/20 Data Partition with Different Models.

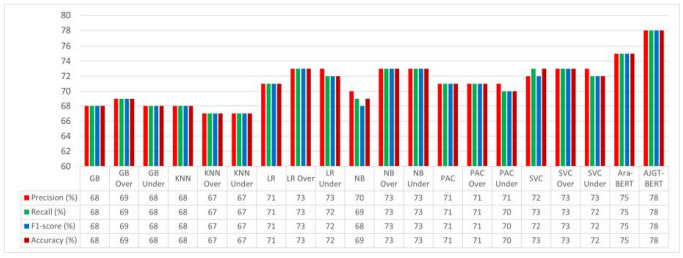

Fig. 4. Various Classification Metrics for 70/30 Data Partition with Different Models.

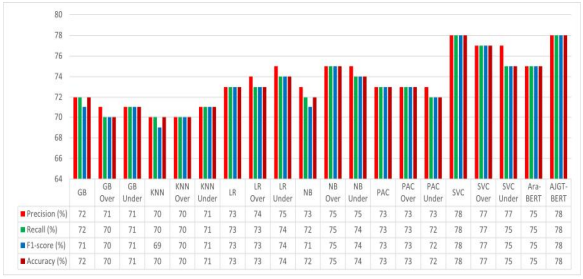

Fig. 5. Various Classification Metrics for 90/10 Data Partition with Different Models.

VI. CONCLUSION

ISSN: 2057-5688

Hate speech is one of the major problems at this time, especially with the increasing number of users on social media. At the same time, an increasing number of crimes became a serious concern that threatens the cohesiveness and structure of civilian societies. Therefore, this work presents an efficient framework to detect Arabic hate speech based on the content of social networks. We utilize various ML and DL models to perform an efficient classification of users' comments. Based on the content of this work, the classes are hate speech or normal. The proposed framework has six ML algorithms and two DL, which are Gradient Boosting, K-Nearest Neighbor, Logistic Regression, Naive Bayes, Passive Aggressive Classifier, Support Vector Machine, Ara-BERT, and BERT-AJGT. For wider investigation, we utilized various scenarios of data partitioning, re-sampling techniques, and transfer learning. We were able to successfully have various classification models with better results in terms of precision, recall, F1-Score, and accuracy compared to the most relevant related work. For future work, we aim to create huge and benchmark data sets. Moreover, working with the mixed language problem, multimodel and data augmentation can be interesting topics for future work on this topic, especially for the Arabic language. In future work, the classification can be expanded to cover many classes such as racism, misogyny, religious discrimination and so on.

REFERENCES

[1] H. Butt, M. R. Raza, M. J. Ramzan, M. J. Ali, and M. Haris, "Attentionbased CNN-RNN Arabic text recognition from natural scene images," Forecasting, vol. 3, no. 3, pp. 520–540, 2021.

[2] R. Duwairi, A. Hayajneh, and M. Quwaider, "A deep learning framework for automatic detection of hate speech embedded in arabic tweets," Arabian Journal for Science and Engineering, vol. 46, no. 4, pp. 4001– 4014, 2021.

[3] I. Guellil, H. Saadane, F. Azouaou, B. Gueni, and D. Nouvel, "Arabic ˆ natural language processing: An overview," Journal of King Saud University-Computer and Information Sciences, vol. 33, no. 5, pp. 497– 507, 2021.

[4] K. Darwish, W. Magdy, and A. Mourad, "Language processing for arabic microblog retrieval," in 21st ACM international conference on Information and knowledge management, 2012, pp. 2427–2430.

[5] J. Chen, Y. Chen, Y. He, Y. Xu, S. Zhao, and Y. Zhang, "A classified feature representation three-way decision model for sentiment analysis," Applied Intelligence, vol. 52, no. 7, pp. 7995–8007, 2022.

[6] G. Kovacs, P. Alonso, and R. Saini, "Challenges of hate speech ´ detection in social media," SN Computer Science, vol. 2, no. 2, pp. 1–15, 2021.

ISSN: 2057-5688

[7] A. Matamoros-Fernandez and J. Farkas, "Racism, hate speech, and ´ social media: A systematic review and critique," Television & New Media, vol. 22, no. 2, pp. 205–224, 2021.

[8] K. Barker and O. Jurasz, "Online misogyny as a hate crime:# timesup," in Misogyny as Hate Crime. Routledge, 2021, pp. 79–98.

[9] A. Al-Hassan and H. Al-Dossari, "Detection of hate speech in social networks: a survey on multilingual corpus," in 6th international conference on computer science and information technology, vol. 10, 2019.

[10] D. Sultan, A. Suliman, A. Toktarova, B. Omarov, S. Mamikov, and G. Beissenova, "Cyberbullying Detection and Prevention: Data Mining in Social Media," in International Conference on Cloud Computing, Data Science & Engineering. IEEE, 2021, pp. 338–342.

[11] S. U. Maheswari and S. Dhenakaran, "Analysis of Approaches for Irony Detection in Tweets for Online Products," in Innovations in Computational Intelligence and Computer Vision. Springer, 2022, pp. 141–151.

[12] I. Kwok and Y. Wang, "Locate the hate: Detecting tweets against blacks," in Twenty-seventh AAAI conference on artificial intelligence, 2013.

[13] A. Y. Muaad, H. J. Davanagere, D. Guru, J. Benifa, C. Chola, H. AlSalman, A. H. Gumaei, and M. A. Al-antari, "Arabic document classification: Performance investigation of preprocessing and representation techniques," Mathematical Problems in Engineering, vol. 2022, 2022.

[14] I. Aljarah, M. Habib, N. Hijazi, H. Faris, R. Qaddoura, B. Hammo, M. Abushariah, and M. Alfawareh, "Intelligent detection of hate speech in arabic social network: A machine learning approach," Journal of Information Science, vol. 47, no. 4, pp. 483–501, 2021.

[15] "Jordanian Ministry of Justice," last accessed September 14, 2022. [Online]. Available: http://www.moj.gov.jo/EchoBusV3.0/SystemAssets/5d38ea27- 5819-443e-a380-b65c7e1f5b56.pdf

[16] M. Masadeh, A. Masadeh, O. Alshorman, F. Khasawneh, and M. Masadeh, "An efficient machine learning-based covid-19 identification utilizing chest x-ray images," IAES International Journal of Artificial Intelligence, pp. 356–366, 2022.

[17] M. Masadeh, O. Hasan, and S. Tahar, "Machine-learning-based selftunable design of approximate computing," IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol. 29, no. 4, pp. 800–813, 2021.

[18] ——, "Machine learning-based self-compensating approximate computing," in

ISSN: 2057-5688

2020 IEEE International Systems Conference (SysCon). IEEE, pp. 1–6.

[19] I. Goodfellow, Y. Bengio, and A. Courville, Deep learning. MIT press, 2016.

[20] R. NL. Stopword lists. [Online]. Available: https://www.ranks.nl/stopwords/arabic

[21] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He, "A Comprehensive Survey on Transfer Learning," Proceedings of the IEEE, vol. 109, no. 1, pp. 43–76, 2021.

[22] R. Mohammed, J. Rawashdeh, and M. Abdullah, "Machine learning with oversampling and undersampling techniques: Overview study and experimental results," in 11th International Conference on Information and Communication Systems (ICICS), 2020, pp. 243–248.