ISSN: 0067-2904

Extremism Detection in the Iraqi Dialect Based on Machine Learning

Redhaa Fadhil Sabri, Nada A. Z. Abdullah*
Department of Computer Science, College of Science, University of Baghdad, Baghdad, Iraq

Received: 1/12/2023 Accepted: 16/2/2024 Published: 28/2/2025

Abstract

Extremism detection is an important area of natural language processing (NLP). It is used to detect hate speech, sectarianism, and terrorism on social media. This field has been discussed and studied in many international languages, especially Arabic and English, as many studies touched on languages in particular, but dialects were not addressed even though users of social networking sites write in their dialect. One of the most difficult Arabic dialects is the Iraqi dialect. Because the Iraqi dialect has few sources on the Internet regarding available data that can be used by researchers, this research aims to detect extremism in Iraqi texts using machine learning. The data was pre-processed by deleting suffixes and prefixes for Iraqi words, deleting repeated letters in the word, and deleting Iraqi stop words. Pre-trained embedding as well as embedding using Gensim Word2vec and FastText were used to represent the words in the embedding step. Also, four learning classifiers were used: Support Vector Machine (SVM), Logistic Regression (LR), K-Nearest Neighbor (KNN), and Gaussian Naive Bayes (GNB). The experiments were conducted on two Iraqi datasets collected from social media platforms related to extremism: the Iraqi Facebook Comments Dataset (IFCD) and the Iraqi Tweets Dataset (ITD). The performance of all models was evaluated using accuracy, macro-average precision, macro-average recall, and macro-average F1-score; the best F1-score is 0.9521, while recall and precision are 0.95 and 0.955, respectively. In addition, the models presented in this research were tested on an Iraqi data set related to hate speech available on the Internet, and the results obtained were compared with the results of the work that provided this data set.

Keywords: Extremism detection, NLP, word embedding, Machine Learning, Iraqi Dialect.

كشف التطرف في اللهجة العراقية باعتماد التعلم الآلي

رضاء فاضل صبري , ندا عبد الزهرة عبدالله*
قسم الحاسوب، كلية العلوم، جامعة بغداد، بغداد، العراق

الخلاصة

يعد كشف التطرف مجالًا مهمًا في معالجة اللغات الطبيعية (*NLP*). يتم استعماله للكشف عن خطاب الكراهية والطائفية والإرهاب على وسائل التواصل الاجتماعي. وقد تمت مناقشة ودراسة هذا المجال في العديد من اللغات العالمية، وخاصة العربية والإنجليزية، كما تطرقت العديد من الدراسات إلى اللغات بشكل خاص،

*Email: Nada.abdullah@sc.uobaghdad.edu.iq

ولكن لم يتم التطرق إلى اللهجات على الرغم من أن مستخدمي مواقع التواصل الاجتماعي يكتبون بلهجاتهم. من أصعب اللهجات العربية هي اللهجة العراقية لأن اللهجة العراقية لديها مصادر قليلة على شبكة الإنترنت فيما يتعلق بالبيانات المتاحة التي يمكن استعمالها من قبل الباحثين، ويهدف هذا البحث إلى الكشف عن التطرف في النصوص العراقية باستعمال التعلم الآلي. تمت معالجة البيانات مسبقاً من خلال حذف اللواحق والبادئات للكلمات العراقية، وحذف الحروف المتكررة في الكلمة، وحذف كلمات التوقف العراقية. تم استعمال التضمين المدرب مسبقًا بالإضافة إلى التضمين باستخدام *Gensim Word2vec* و*FastText* لتمثيل الكلمات في خطوة التضمين. كما تم استعمال أربع مصنفات تعلم آلي: *Support Vector Machine* (*SVM*)، *Logistic Regression* (*LR*)، *Knearest neighbor* (*KNN*)، و *Gaussian Naïve Bayes (GNB)*. أجريت التجارب على مجموعتي بيانات عراقية تم جمعهما من منصات التواصل الاجتماعي تخص التطرف: مجموعة بيانات تعليقات فيسبوك عراقية (*IFCD*) ومجموعة بيانات تغريدات عراقية (*ITD*). تم تقييم أداء جميع النماذج باستعمال الدقة، والدقة المتوسطة الكلية، والاستدعاء المتوسط الكلي، والمتوسط الكلي *F1-score* افضل دقة هي *0 .9521* والدقة المتوسطة الكلية، والاستدعاء المتوسط *0.95* و*0.955* بالتتابع. بالإضافة إلى ذلك، تم اختبار النماذج المقدمة في هذا البحث على مجموعة بيانات عراقية تتعلق بخطاب الكراهية متوفرة على الإنترنت، وتمت مقارنة النتائج التي تم الحصول عليها مع نتائج العمل الذي قدم مجموعة البيانات هذه.

1. Introduction

Because the Internet, especially social networking sites, is full of users of different nationalities and religions who are now able to freely express their opinions on matters of life regarding religion, politics, and various other interests, publishing their opinions and broadcasting their thoughts through these platforms has become easy. And with the click of a button, their opinions become published on the Internet to be seen by other users, who also express their thoughts and opinions. The fact that the number of users of these programs is constantly increasing makes it impossible to control and review all publications to see if they contain hate speech or insults to any party, person, or religion. This helped in the spread of extremism among users on these platforms quite easily, especially in the Arab world and Iraq in particular, given that it was and still is suffering from the effects of wars, sectarianism, and internal and external conflicts, so members of society began to express their opinion on their government's policy, their religions, and various matters on social media platforms, especially on Twitter and Facebook (Meta). Hate speech is one of the global problems for which it is difficult to give a unified definition due to the differences in cultures, customs, and traditions from one society to another. In addition, this field suffers from a lack of research studies targeting it on online social networks, which has caused great challenges for researchers [1]. Most of these opinions are those of fanatics that contain extremism or hate speech. Many researchers have presented a lot of studies and research related to extremism and hate speech in Modern Standard Arabic (MSA), as well as studies in most Arabic dialects, but the studies that were in the Iraqi dialect and related to extremism were unfortunately very few compared to the rest. Most of the data on these platforms contains noise that increases the spacing between words, and there are no appropriate pre-processing steps for the data, especially those written in the Iraqi dialect. There was also a problem: the use of pre-trained word embedding models as feature extraction, which helps in classifying and analyzing hate speech and extremist content with data written in the Iraqi dialect, led to the emergence of out-of-vocabulary problems because these models are specific to Modern Standard Arabic. Therefore, the research aim is to encourage the study of this field in the Iraqi dialect because the number of users of social networking sites is very large compared to other users, and these users express their opinions and ideas in their dialect in their publications. It also aims to encourage the development of methods of dealing with the Iraqi

dialect because it is one of the most difficult and widespread Arabic dialects. Therefore, we hope in the future to notice a significant development regarding the Iraqi dialect in the field of natural language processing (NLP).
The main contributions in this paper are as follows:

- Create new preprocessing steps to deal with prefixes and suffixes in Iraqi words (stem).
- Create Iraqi embedding vectors.
- Build a machine learning model specialized to classify extremist text and hate speech in the Iraqi dialect.
- Make the classification process more accurate by utilizing extreme emojis.

The remainder of the paper will have the following: The section explores a review of extremism detection-related work in Modern Standard Arabic (MSA) and Arabic Dialects; the third section shows the research methodology; the fourth section will show and discuss the results of the proposed models; the fifth section will compare the results obtained by the four proposed models with the results of the related work; and the conclusion and future work will be in the last section.

2. Related Work

Bayan M. Sabbar et al. (2018) [2] collected data in the Iraqi dialect related to politics and the Iraqi government from the comments of Iraqi citizens on Facebook and then applied a set of pre-processing steps to it. One of the most important of these steps is to build a stem for the Iraqi dialect to reduce vocabulary as much as possible. This was the first attempt to build a stem specialized in the Iraqi dialect. Then they applied two approaches to the data that were processed, which are the machine-learning approach and the lexicon-based approach. In the machine learning approach, three algorithms were applied, which are Naive Base (NB), K-Nearest Neighbor (KNN), and AdaBoost Ensemble. The highest result was obtained through NB, with a multinomial of 93%. While the research approach has its advantages, some weaknesses must be considered in the future. Researchers relying on a single data set may lead to bias and limit the diversity of destinations. The effectiveness of pre-processing steps also depends on the quality of the algorithms and tools used. The stem developed for the Iraqi dialect may not cover all linguistic differences.

According to Anwar Alnawas et al. 2019 [3], the Iraqi dialect was addressed in particular through sentiment analysis of the Iraqi comments that were collected from Facebook, and then word embedding was used to create a large corpus to represent the data first. Secondly, they created a word embedding model by training the corpus that was created first using Doc2Vec, then four machine learning classifiers were applied to it, such as Logistic Regression (LR), Decision Tree (DT), SVM, and NB, and the highest score they obtained was 82% through SVM. A single data set was used, which means that it could lead to bias and a lack of diversity, which affects the validity of the results because it does not cover the largest possible difference in dialect. In addition to building a word embedding for a small amount of data that does not cover a variety of linguistic differences, the resulting embedding may not effectively reflect the richness and complexity of the Iraqi dialect.

Mohammed A. Alghamdi et al. 2020 [4], in this study, introduced an intelligent system for analyzing Arabic tweets to detect suspicious messages. Data was collected from Twitter using the Twitter API and saved in a file in a specific format. The proposed system encoded and processed the data, and then manual labels were applied to it, as the suspicious tweets were (1) and non-suspect (0). After this step, six machine learning algorithms were applied to it, such as (1) DT, (2) KNN, (3) linear discriminant algorithm (LDA), (4) SVM, (5) artificial neural networks (ANN), and (6) long short-term memory networks (LSTM). The highest accuracy obtained is 86.72% through SVM. Manual labeling of data is a very important process that increases the accuracy of classification. It requires a lot of time and effort, in addition to the fact that some data can cause conflicts and make it difficult for researchers to classify it manually.

Ahmed I. A. Abd-Elaal et al. 2020 [5] proposed a smart system that detects the ISIS community on the Twitter platform. This system analyzes hashtags related to ISIS and monitors the accounts of people interested in these hashtags, as well as monitoring mentions. This system consists of two sub-systems: the crawling system and the inquiry sub-system. Term frequency-inverse document frequency (TF-IDF) and Skip-gram as feature extraction, and then the proposed system was tested by some machine learning algorithms, and the results showed the highest accuracy of 94% through the linear SVM algorithm Skip-gram embedding. Collecting data on the ISIS community and monitoring hashtags related to them may lead to bias because it is assumed that individuals using these hashtags are necessarily associated with the ISIS community. Also, the focus on discovering the ISIS community may limit the possibility of generalizing the proposed system to other extremist groups.

Rawan Abdullah Alraddadi et al. (2021) [6] proposed a classification of anti-Islamic Arabic texts using text mining and sentiment analysis techniques. A set of appropriate data was collected for this topic, and pre-processing steps were applied to it. TF-IDF was then used for feature extraction, and then two machine learning algorithms were applied to the data, such as SVM and Multinomial Naive Bayes (MNB). The highest accuracy was 97% obtained through SVM. Using TF-IDF to extract features is common, but it may not capture the semantic relationships between words. The quality and representativeness of the data collected are critical to the generalizability of the model. If the data set does not cover a wide range of contexts, the effectiveness of the model may be limited.

Mohammed M. Hassoun Al-Jawad et al. 2022 [7]: This study presented an Iraqi data set related to extremism and hate speech that was collected from Twitter from Iraqi hashtags that are related to political matters and the Iraqi government. It consisted of 1170 tweets, some of which contained hate speech and others did not contain hate speech, collected through the Twitter application programming interface (API). The researchers made the data available to everyone to encourage the creation of studies on this subject in the Iraqi dialect. Then they applied some machine learning classifiers to the data to classify whether it contained hate speech or not. They obtained the highest accuracy of 78% through LibSVM. The size of the data is relatively small, especially for machine learning models. In addition, the small number of data points causes significant bias because they do not cover the diversity of the data. Since the data concerns hate speech, a balance must be taken into account between the number of texts that contain hate speech and those that do not.

Khalid T. Mursi et al. (2022) [8] presented a group of manually classified tweets, which consisted of 3,000 Arabic tweets. This data contains tweets that contain hate speech and tweets that do not contain hate speech. To classify these tweets, Word2Vec word embedding was used, and they also applied their proposed model to 100,000 tweets from the past decade. The proposed model achieved an accuracy of 92% through SVM. The number of data points is relatively small. In addition, the study must provide details about the extent to which the manually classified data represents the broader population of Arab Twitter users. If the training data is not sufficiently diverse, the model may not perform well on tweets from different sources or contexts. Focusing on the Arabic language weakens the possibility of generalizing the model to Arabic dialects.

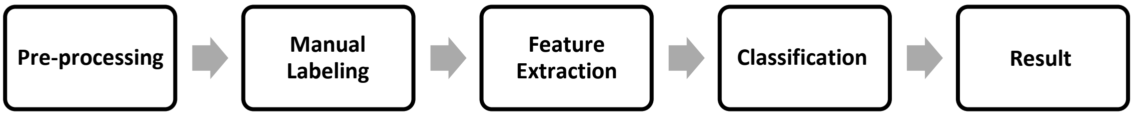

Figure 1: Methodology of the Research

3. Research Methodology

As we mentioned earlier, there is very little research on extremism and hate speech in the Iraqi dialect. This has led to the fact that the data in the Iraqi dialect available on the Internet regarding extremism that was collected from social networking sites is very limited compared to data in other dialects and the Arabic language. In addition to the fact that the data collection process on social networking sites requires you to obtain permission and approval from those platforms, using programs dedicated to data collection takes a lot of time. The three datasets used in this research related to extremism were collected from the Iraqi social media sites Facebook (Meta) and Twitter (X). An Iraqi Tweets dataset (ITD) was collected from Iraqi hashtags on Twitter, consisting of 3100 tweets written in the Iraqi dialect. The second dataset was collected from an Iraqi page on Facebook, consisting of 12,000 comments written in Iraqi dialect. Finally, the third dataset that used CIAD was collected from Iraqi hashtags on Twitter written in the Iraqi dialect, consisting of 1170 tweets available on the internet [7].

3.1 Preprocessing

The preprocessing process is important when dealing with text data to identify features that represent the document semantically and remove features that do not. Therefore, the primary goal of this stage is to reduce the error rate and test space [9]. The Iraqi dialect is considered one of the most difficult dialects because the Arabic language contains 28 letters written from right to left, but the Iraqi dialect, in particular, has many additions, such as the letter (چ), which is pronounced (ch), and sometimes it is written (ج), and (گ), which is pronounced (ga) and written (ك), but both letters are pronounced as we mentioned earlier. In addition, the Iraqi dialect contains many words that do not exist in the Arabic language, and therefore you need to take the corresponding words in the Arabic language, i.e., their synonyms, to facilitate work with this data as well. Since the data is collected from social platforms, this means that there are many spelling errors in addition to shortening words or encoding some speech, and all of the above causes difficulty and challenges in dealing with data. Therefore, in this research, we focused mainly on the preprocessing. Below, we will explain the preprocessing in detail.

3.1.1 Clean Data

In this step, we remove:

- URL link
- User name
- Symbols such as (#, @, !, $, %, ^, &, *….)
- Remove useless emojis and convert all emojis that indicate offense to the word (Extremism) such as (💣, , ……)
- Normalize letters such as (چ, گ, أ, آ, ئ, ة, ؤ)

- Remove English and Arabic letters and numbers

3.2.2 Tokens

It is the division of speech into a group of words, and each word is called a token [10]. In this research, each tweet was divided into a group of words to facilitate processing.

3.2.3 Remove Iraqi Stop words

Stop words in general mean those words that, if we remove them, do not affect the context and meaning of speech, such as prepositions, pronouns, question articles, and others. These words, such as conjunctions, pronouns, and others, have no effect, and usually, these words add very little meaning and may not be added, so when they are removed from the text, it does not affect the meaning of the text [11].

3.2.4 Remove Single Letters

Some users abbreviate words to one letter, such as (on, على) to (ع) and other abbreviations and single letters that have no important meaning.

3.2.5 Iraqi Dialect Stem

The stemming process is the process of returning the word to its root. Also, roots are connected to the word either at the beginning (called the prefix), in the middle (called the infix), or at the end (called the suffix) [12]. Since the Iraqi dialect is somewhat different from modern standard Arabic, this process is somewhat different. There are two studies in which the researchers touched on building a stem for the Iraqi dialect.

The researchers in this study created a set of steps for a stem specialized in the Iraqi dialect, where they processed the words by deleting the prefixes such as (يال، كال فال، وال، بال، ولل، لل، ال، و) where they deleted the additions at the beginning of words to reduce the vocabulary as well as make it close to the Arabic language and other dialects and be more understandable, depending on the length of the word [2].

The researchers built steps for a stem that also specializes in the Iraqi dialect by deleting the prefixes at the beginning of words such as (و ، بهال، بال، ب، هال شد، ال، ش، حي، د), and these additions come in most words written in the Iraqi dialect; omitting them helps to bring the words closer to the Arabic language [13].

As for our research, we took advantage of both studies in making a stem specialized in the Iraqi dialect, but we added some steps that were not addressed in both studies, as we dealt with prefixes as well as suffixes that come at the beginning and end of words. Prefixes such as (م، يال، كال، فال، وال، بال، ولل، لل، ال، و، بهال، ب، هال، شد، ش، حي، د) and suffixes such as (ج، كم، هم، لكم، لهم، ون، جن، وا) but taking into account when deleting whether the letter is original within the words and when deleting it affects the meaning of the word or not, by creating a text file that contains all the words in which these letters are original and cannot be deleted. Some examples of our Iraqi stem are in Table 1, and the letters that are written in red are the prefixes and suffixes in the Iraqi dialect.

Table 1: Examples of Iraqi stem

Words before stem	Words after stem
اعرفكم شرفاء ليش متغردون على صعود الدولار وين زميلك حيدر البرزنجي	اعرف شرفاء ليش تغرد علي صعود الدولار وين زميلك حيدر البرزنجي
تره صارلكم سنه تحكمون يعني شكد لازم تبقون حتي تكولون احنا فاشلين	تره صار سنه تحكم يعني شكد لازم تبق حتي تكول احنا فاشلين
يارب اوصل هلمرحلة بالرسم مثلج تخبل كلش	يارب اوصل مرحلة بالرسم مثل تخبل كلش

3.2 Manual Labeling

The text in the three datasets was manually labeled into two categories: offensive (1) and non-offensive (0). This is because our work is limited only to classifying the tweets to determine whether they contain extremism, that is, whether the type of extremism is present or not, and this classification is called the binary classification. The manual labeling process depends entirely on the person who performs it, as human judgment and intelligence play a role in the labeling process. This study summarized [14] that the classification performance depends more on the size and quality of the training data than its type. However, the manual labeling process needs to be carried out by the machine because it is hard work and takes a lot of time, especially if the size of data used is very large, but taking into account the human verification of the labeling because the machine sometimes does not understand whether the words are intended as an insult or not as an example (هذا واحد كلب وحيوان) Here, this sentence is intended as an insult, but the machine can consider it non-offensive. The manual labeling of data, although it requires more time and requires that the person who performs it be well aware that the data contains extremism or not, was accurate, and most of the studies that it conducted obtained very high results. This proves that this method is very effective, especially if the number of data points is small, and the most important of these studies that have been carried out in this way are [5], [15], [16], [8], [17], etc.

3.3 Feature Extraction

The process of extracting the features that represent the text in a format that the machine can understand through the numerical representation of words, that is, replacing the text with numbers, because the machine deals with the language of numbers only, as well as machine learning classifiers that deal with numbers, There are many techniques used for feature extraction in natural language processing. One of the most important of these techniques is word embedding, which converts words into vectors of numbers and is a way of mapping words into a high-dimensional vector space [18]. One of the most important feature extraction methods is word embedding, which is done through several methods such as Word2Vec, which have been used in [5], [8], and [19]. Also, there are pre-trained models such as Glove [20] and FastText. In this paper, we use FastText, which is developed by Facebook (Meta) [21]. A pre-trained model for the Arabic language, which can be found and downloaded through FastText's website, and a Glove pre-trained model were used. Also, Iraqi word embedding models were built using FastText and Word2Vec.

3.4 Classification

There are many machine learning algorithms, some of which are called supervised learning, unsupervised learning, and reinforcement learning [22], but since our research aims to classify tweets into a binary classification and we also made manual labeling of our data set, the best algorithms that fit our work are supervised learning algorithms such as Logistic Regression, Gaussian Naive Bays, K-Nearest Neighbors, and Support Vector Machine. Below, we will give a simplified explanation for each one:

3.4.1 Logistic Regression (LR):

It is a linear model that estimates the probability of a variable belonging to a class and is used in binary classification problems. It is also used in many areas, including spam email classification and fraud detection.

3.4.2 Gaussian naïve Bayes (GNB):

The probabilistic learning algorithm is usually used with NLP problems and the classification of textual data. calculates the probability of an event occurring based on prior knowledge of the circumstances related to that event. This algorithm is one of the simplest and easiest to implement, and it can easily handle large data sets. Gaussian Naive Bayes (or normal distribution) is the easiest to work with because you only need to estimate the mean and standard deviation from your training data [23].

3.4.3 K-Nearest Neighbors (KNN):

This algorithm is a supervised algorithm that can be used for classification [24]. KNN works by taking advantage of the similarities between the examples (tweets). The dataset is for training tweets with labels, in our case, suspicious tweets (1) and non-suspicious tweets (0). If you are given a tweet without a label from the data test, it compares its vocabulary with that of the trained tweets. then take the most similar tweets and search for the most similar tweets from the trained tweets. Usually, the value of K is less than 20.

3.4.4 Support Vector Machine (SVM):

SVM is widely used to solve classification problems in many areas, the most important of which are text data and images. The algorithm creates training where it builds a model that assigns examples (the text of the tweets) to categories (suspicious and not suspicious) [25]. The SVM aims to produce a tweet classification model based on the set of tweets you are training. The Sequential Minimum Optimization (SMO) algorithm can be used to train support vector machines.

Several machine learning algorithms were used on the three datasets after they were converted into vectors during the feature extraction process. KNN, SVM, GNB, and LR were used. The four proposed models are shown in Figure 2, Figure 3, Figure 4, and Figure 5.

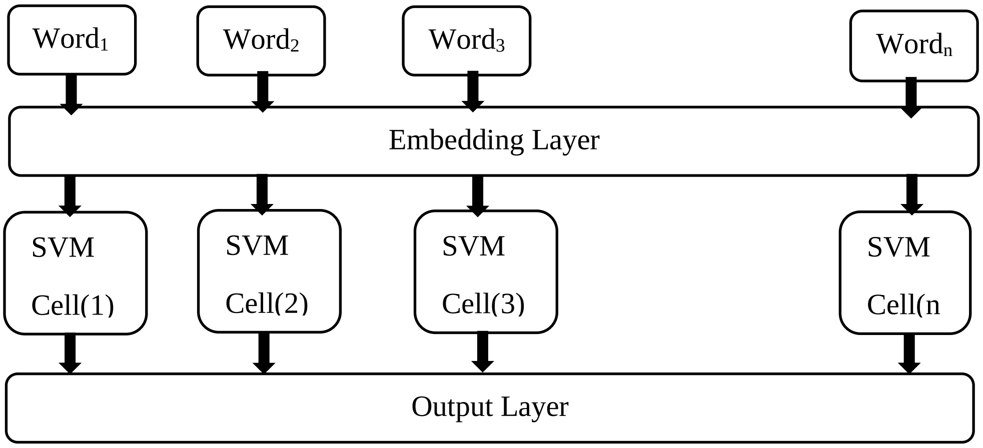

Figure 2: Proposed SVM Extremism Detection Model

Figure 2 shows the layers of the proposed SVM model, which consists of four layers: the input, embedding layer, SVM layer, and the output layer, which we will explain below.

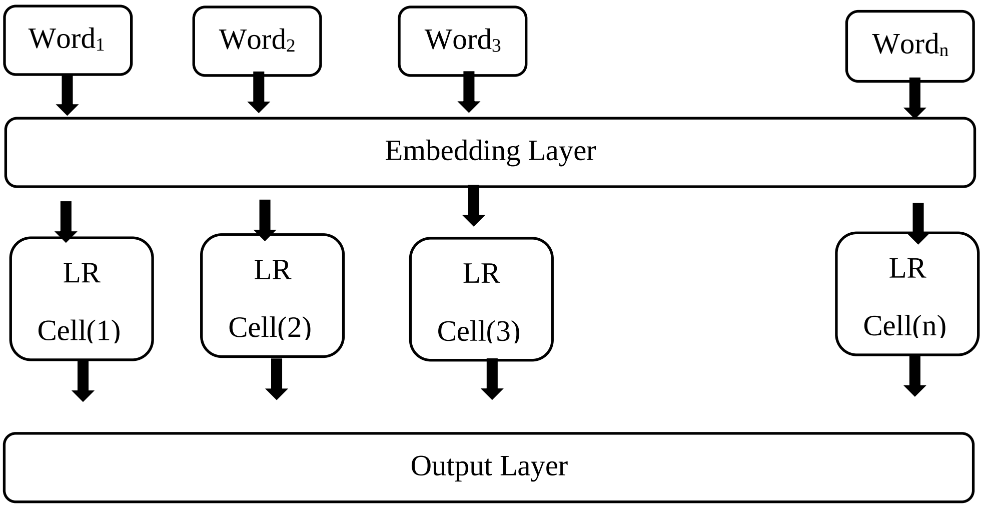

Figure 3: Proposed LR Extremism Detection Model

Figure 3 shows the layers of the proposed LR model, which also consists of four layers: the input, embedding layer, LR layer, and output layer.

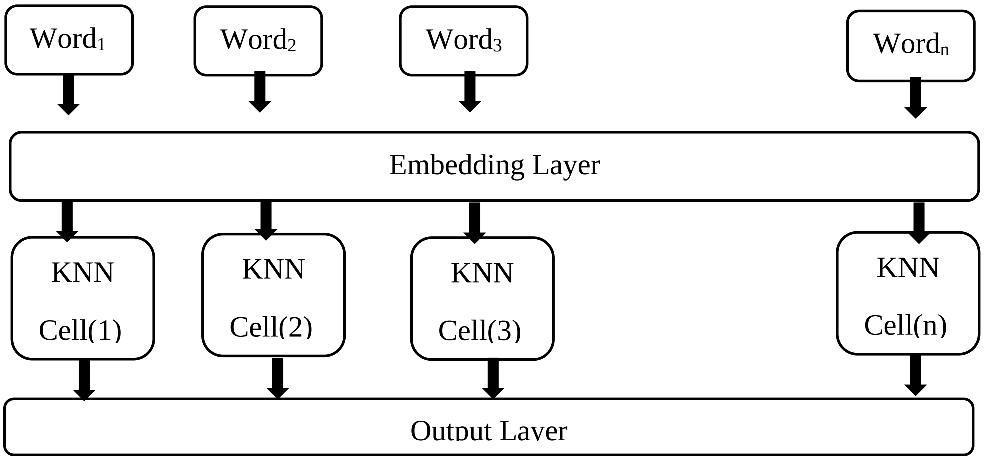

Figure 4: Proposed KNN Extremism Detection Model

Figure 4 shows the layers of the proposed KNN model, which consists of four layers represented by the input, embedding layer, KNN layer, and output layer.

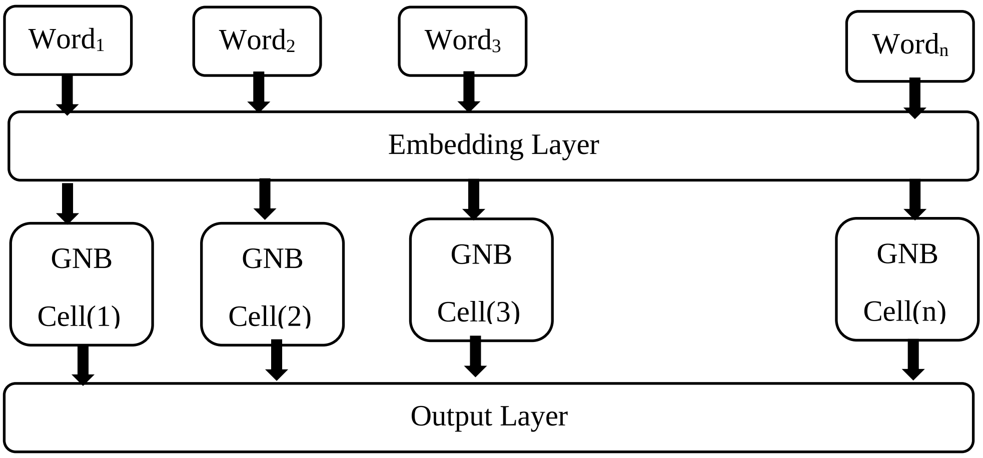

Figure 5: Proposed GNB Extremism Detection Model

Figure 5 shows the layers of the proposed GNB model, which consists of four layers represented by the input, embedding layer, KNN layer, and output layer.
The four layers of all proposed models will be explained below:

1. *The Input Layer:*

It is the first layer in all proposed models, and in this layer, the input is text split into sentences, and each sentence is split into words.

2. *The Embedding Layer:*

In this layer, all texts are represented as an embedding matrix. The matrix of this layer is ***w*v***, where ***w*** is the maximum number of words in the text and ***v*** is the vector length to represent the word. Several embedding techniques were used, such as FastText, Word2Vec, and Glove, as mentioned in the feature extraction section.

3. *The classifier layer:*

This layer takes the output of the previous layer as input. We used four ML classifiers (SVM, LR, KNN, and GNB). All classifiers will take the output of the previous layer as an input to find a decision boundary that best separates the classes in the feature space, leading to accurate text classification.

4. *The Output Layer:*

Depending on how many classes need to be identified, the final layer is added to transform the input vector from the previous layer into a single output. Two classes were used because the classification is binary. The output of this layer is 1 if the text has extremism or 0 if it doesn't.
The algorithm of the proposed models is shown below. All algorithms are the same for all models, taking into account classifier variability.

Algorithm 1: The proposed model algorithm.

```
Input: Features vectors (with different lengths), The embedding matrix.
Output: Classification results, Confusion matrix.

Begin:
  Step 1: Create the first layer by using the embedding matrix.
  Step 2: Select features from the previous layer using the Average pooling function.
  Step 3: Split the pooled features matrix into training and testing sets.
  Step 4: Train the four classifiers using training features and corresponding labels.
  Step 5: Use the trained classifiers to make predictions on the testing features.
  Step 6: Evaluate the classification results using metrics (accuracy, Macro_Average
  precision, Macro_Average recall, Macro_Average F1-score).
  Step 7: Return the classification results.
End
```

4. Results and Discussion

The Google Colab notebook environment was used to implement the system's relatively inexpensive memory, longer runtimes, and faster GPUs. The RAM was around 12.7 GB, and the disk utilized to run the model libraries was 107.72 GB in size. Python 3.10 was used as a programming language with libraries such as sklearn (1.2.2), Numpy (1.23.5), Gensim (4.3.1), Panadas (1.5.3), and Keras (2.12.0). The Central Processing Unit (CPU) Colab was an Intel (R) Xeon (R) processor running at 2.20 GHz, and the Graphics Processing Unit (GPU) was an NVIDIA Tesla K80 GB. Metrics such as accuracy, macro-averaged precision, macro-averaged recall, and macro-averaged F-score are used to evaluate the two datasets, which are calculated in Eqs. 1 to 4. The highest results were obtained by the SVM extremism detection model and an Iraqi Word2vec model for feature extraction with an accuracy of 96%. The results of all classification models are shown in Table 2, with the highest accuracy in bold, and the results of other models are shown in Table 3, Table 4, and Table 5.

$$Accuracy = \frac{(TP + TN)}{(TP + FN + FP + TN)} \tag{1}$$

$$Precision = \frac{TP}{(TP + FP)} \tag{2}$$

$$Recall = \frac{TP}{(TP + FN)} \tag{3}$$

$$F - score = \frac{\left((\beta^2+1)TP\right)}{\left((\beta^2+1)TP+\beta^2FN+FP\right)} \tag{4}$$

With binary classification, the most widely used performance measures based on a matrix called the confusion matrix (CM) are accuracy (AUC), precision, recall, and f-score, and these measures are widely used with texts [24].

Also, the CM called the error matrix, based on the actual and predicted results of a set of testing data, is used to evaluate the performance of the model in the supervised classifiers, and it has four values: TP, TN, FP, and FN [26].

These values mean, as follows [27]:

- False negative (FN): The number of incorrectly labeled negative texts.
- False positive (FP): The number of incorrectly labeled positive texts.
- True positive (TP): The number of correctly labeled positive texts.
- True negative (TN): The number of correctly labeled negative texts.

Table 2: Result of Classification with the Iraqi Word2vec Model as Feature Extraction

Classifiers	Macro avg			Accuracy
	Precision	**Recall**	**F1-score**	
LR	0.93	0.94	0.93	0.94
GNB	0.79	0.80	0.79	0.80
KNN	0.93	0.91	0.92	0.92
SVM	0.96	0.95	0.95	**0.96**

Table 2 shows that the best performance was achieved by the SVM model as a classifier and Iraqi Word2vec as a feature extraction with an accuracy of 0.96.

Table 3: Result of Classification with the Iraqi FastText Model as Feature Extraction

Classifiers	Macro avg			Accuracy
	Precision	**Recall**	**F1-score**	
LR	0.89	0.88	0.88	0.89
GNB	0.79	0.80	0.79	0.80
KNN	0.92	0.90	0.92	0.92
SVM	0.94	0.93	0.93	**0.94**

Table 3 shows that the best performance was achieved by the SVM model as a classifier and Iraqi FastText as a feature extraction, with an accuracy of 0.94.

Table 4: Result of Classification with the FastText Pre-trained Model

Classifiers	Macro avg			Accuracy
	Precision	Recall	F1-score	
LR	0.76	0.76	0.76	0.76
GNB	0.70	0.70	0.70	0.71
KNN	0.92	0.91	0.91	0.91
SVM	0.94	0.90	0.92	**0.93**

Table 4 shows that the best performance was achieved by the SVM model as a classifier and the FastText pre-trained model as a feature extraction, with an accuracy of 0.93.

Table 5: Result of Classification with a Glove-Pretrained Model

Classifiers	Macro avg			Accuracy
	Precision	Recall	F1-score	
LR	0.90	0.87	0.88	0.89
GNB	0.77	0.77	0.76	0.77
KNN	0.92	0.91	0.91	0.91
SVM	0.95	0.93	0.94	**0.95**

Table 4 shows that the best performance was achieved by the SVM model as a classifier and the Glove pre-trained model as a feature extraction, with an accuracy of 0.95.
All experimental results presented in the previous tables show that the SVM model achieved the highest accuracy with the four-word embedding models. That means SVM is a more accurate classifier with text data.

5. Comparing the Proposed Models' Results with other Related Work

In this section, all models proposed are applied to the dataset CIAD provided by [7]. Table 6 shows the results of the proposed models and their results.

Table 6: Comparison between proposed model results and related work

Related Work	Classifier	Accuracy	Macro-Average		
			F1-score	Recall	precision
Our proposed models	SVM	**0.9554**	**0.9521**	**0.9500**	**0.9545**
	LR	0.9375	0.9336	0.9357	0.9317
	KNN	0.8929	0.8805	0.8667	0.9065
	GNB	0.8036	0.7924	0.7952	0.7901
Mohammed M. Hassoun Al-Jawad et al. **2022 [7]**	SMO	76.2	81.2	63.7	71.4
	LibSVM	78.1	81.2	68.9	74.5

They used two versions of SVM on the dataset, which are Sequential Minimal Optimization (SMO) and Library for Support Vector Machine (LibSVM), and the highest accuracy they obtained was 78.1% by LibSVM and 76.2% by SMO. The proposed model's highest accuracy is 0.96 by SVM, 0.94 by LR, 0.89 by KNN, and 0.80 by GNB. The table above shows that the outperformance of all proposed models achieved the highest scores in terms of macro-average f1-score, macro-average recall, macro-average precision, and accuracy.

6. Conclusion and Future Work

After we discussed the results of the proposed models, we concluded the following:
1- Building the Iraqi vector model gave the best results compared with pre-trained models.

2- The pre-processing steps that were performed on the data led to clear words and made the classification process more accurate.
3- The process of manual labeling before the embedding stage led to good results and also reduced the excess supply.
4- The Iraqi Word2vec model gave the best accuracy with the support vector machine classifier.
5- The proposed Iraqi stemmer makes the vocabulary clearer and closer to Arabic words.
We hope in the future to build models for pre-trained Iraqi vectors on large data sets to benefit from them in the field of natural language processing and make them available to the public, similar to the pre-trained models for other Arabic dialects. We also hope that a specialized stem will be built in the Iraqi dialect and be unified and standard to encourage the Iraqi dialect to be used and conduct research on it by researchers in the field of language processing because it is one of the most important and most used dialects on the Internet today. Also, the datasets related to extremism written in the Iraqi dialect are very poor on the internet; we need to create more Iraqi datasets to encourage researchers in this field.

References

[1] 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. 3, pp. 1-19, 2020.

[2] B. M. Sabbar, N. T. Yousir and L. A. Habeeb, "Sentiment Analysis For Iraqis Dialect In Social Media Using Machine Learning Algorithms," *Iraqi Journal of Information and Communications Technology(IJICT),* vol. 1, no. 2, pp. 24-32, 2018.

[3] A. Alnawas and N. Arici, "Sentiment Analysis of Iraqi Arabic Dialect on Facebook Based on Distributed Representations of Documents," *ACM Trans. Asian Low-Resour. Lang. Inf. Process,* vol. 18, no. 3, pp. 1-17, 2019.

[4] M. A. AlGhamdi and M. A. Khan, "Intelligent Analysis of Arabic Tweets for Detection of Suspicious Messages," *Arabian Journal for Science and Engineering,* vol. 45, no. 8, pp. 6021-6032, 27 2 2020.

[5] A. I. Abd-Elaal, A. Z. Badr and H. . M. K. Mahdi, "Detecting Violent Radical Accounts on Twitter," *(IJACSA) International Journal of Advanced Computer Science and Applications,* vol. 11, no. 8, pp. 516-522, 2020.

[6] R. A. Alraddadi and M. I. El-Khalil Ghembaza, "Anti-Islamic Arabic Text Categorization using Text mining and Sentiment Analysis Techniques," *(IJACSA) International Journal of Advanced Computer Science and Applications,* vol. 12, no. 8, pp. 776-785, 2021.

[7] M. M. H. Al-Jawad, H. Alharbi, A. F. Almukhtar and A. A. Alnawas, "Constructing twitter corpus of Iraqi Arabic Dialect (CIAD) for sentiment analysis," *Scientific and Technical Journal of Information Technologies, Mechanics,* vol. 22, no. 2, p. 308–316, 2022.

[8] K. T. Mursi, M. D. Alahmadi, F. S. Alsubaei and A. S. Alghamdi, "Detecting Islamic Radicalism Arabic Tweets Using Natural Language Processing," *IEEE Access,* vol. 10, pp. 72526-72534, 2022.

[9] F. A. Abdulghani and N. A. Abdullah, “A Survey on Arabic Text Classification Using Deep and Machine Learning Algorithms,” *Iraqi Journal of Science*, vol. 63, no. 1, pp. 409–419, Jan. 2022.

[10] M. A. H. Wadud, M. F. Mridha, and M. M. Rahman, “Word Embedding Methods for Word Representation in Deep Learning for Natural Language Processing,” *Iraqi Journal of Science,* vol. 63, no. 3, pp. 1349–1361, Mar. 2022.

[11] A. I. Kadhim, "An Evaluation of Preprocessing Techniques for Text Classification," *International Journal of Computer Science and Information Security (IJCSIS),* vol. 16, 2018.

[12] R. A. Sameer, "Modified Light Stemming Algorithm for Arabic Language," *Iraqi Journal of Science,* vol. 57, no. 1B, pp. 507-513, Feb. 2023.

[13] T. Z. Abdulhameed, I. Zitouni and I. Abdel-Qader, "Wasf-Vec: Topology-based Word Embedding for Modern," *ACM Transactions on Asian and Low-Resource Language Information Processing,* vol. 19, no. 2, p. 1–27, 2019.

[14] I. Mozetič, M. Grčar and J. Smailović , "Multilingual Twitter Sentiment Classification: The Role of Human Annotators," *PLoS ONE,* vol. 11, no. 5, pp. 1-26, 2016.

[15] W. Sharif, S. Mumtaz, Z. Shafiq, O. Riaz, T. Ali, M. Husnain and G. S. Choi, "An Empirical Approach for Extreme Behavior Identification through Tweets Using Machine Learning," *Applied Sciences,* vol. 9, no. 18, pp. 1-20, 2019.

[16] N. Albadi, M. Kurdi and S. Mishra, "Are They Our Brothers? Analysis and Detection of Religious Hate Speech in the Arabic Twittersphere," *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM),* pp. 69-76, 2018.

[17] M. Ashcroft, A. Fisher, L. Kaati, E. Omer and N. Prucha, "Detecting Jihadist Messages on Twitter," *2015 European Intelligence and Security Informatics Conference,* pp. 161-164, 2015.

[18] A. A.-R. Alfarhany and N. A. Z. Abdullah, "Iraqi Sentiment and Emotion Analysis Using Deep Learning," *Journal of Engineering,* vol. 29, pp. 150-165, 2023.

[19] S. Aldera, A. Emam, M. AL-qurishi, M. Alrubaian and A. Alothaim, "Exploratory Data Analysis and Classification of a New Arabic Online Extremism Dataset," *IEEE Access,* vol. 9, pp. 161613-161626, 2021.

[20] J. Pennington, R. Socher and C. D. Manning, "GloVe: Global Vectors for Word Representation," *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP),* vol. 14, p. 1532–1543, 2014.

[21] P. Bojanowski, E. Grave, A. Joulin and T. Mikolov, "Enriching Word Vectors with Subword Information," *Transactions of the Association for Computational Linguistics,* vol. 5, no. 1, pp. 135-146, 2017.

[22] T. Kanan, O. Sadaqa , A. Aldajeh , H. Alshwabka, W. AL-dolime , S. AlZu'bi, M. Elbes , B. Hawashin and M. A. Alia , "A Review of Natural Language Processing and Machine Learning Tools Used to Analyze Arabic Social Media," *2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT),* pp. 622-628, 2019.

[23] B. Mahesh, "Machine Learning Algorithms - A Review," *International Journal of Science and Research (IJSR),* vol. 9, no. 1, pp. 381-386, 2020.

[24] B. T.K., C. S. R. Annavarapu and A. Bablani, "Machine learning algorithms for social media analysis: A survey," *Computer Science Review,* vol. 40, pp. 1-32, 2021.

[25] R. Agrawal, M. Paprzycki and N. Gupta, Big Data, IoT, and Machine Learning Tools and Applications, first ed., New York: Library of Congress, 2021.

[26] A. Kumar and A. Jaiswal, "Systematic literature review of sentiment analysis on Twitter using soft computing techniques," *Concurrency and Computation Practice and Experience,* vol. 32, no. 4, pp. 1-29, 2019.

[27] G. A. Ruz, P. A. Henríquez and A. Mascareño, "Sentiment analysis of Twitter data during critical events through Bayesian," *Future Generation Computer Systems,* vol. 106, pp. 92-104, 2020.