Mansoura Journal for Computer and Information Sciences (2024), Vol.**19**, No.**1**
doi:10.21608/mjcis.2024.275844.1002

Mansoura Journal for Computer and Information Sciences

ARTICLE

XAI-Based Sentiment Analysis Using Machine Learning Approaches

Ahmed Elbasiony,* Ibrahim M. El-Hasnony, and Samir Abdelrazek

Department of Information Systems, Faculty of Computers and Information, Mansoura University, Mansoura, 35516, Egypt
*Corresponding author: ahmed.basuony2389@gmail.com

(Received: 10 March 2024; Accepted: 3 September 2024; Published: 7 September 2024)

Abstract
Sentiment analysis is essential for comprehending public views on various issues. While Sentiment Analysis (SA) techniques have been widely adopted, the lack of transparency in conventional machine learning models inhibits a comprehensive understanding of the reasoning behind sentiment predictions, where this opacity hinders the trustworthiness of sentiment analysis models, limiting their applicability in real-world scenarios where interpretability is crucial. It is even worse when sentiment analysis is applied to Arabic text since the language's intricacy and cultural quirks produce particular difficulties, so the problem addressed in this paper revolves around the need for accurate sentiment analysis in Arabic textual data and providing the feature of interpreting the results of sentiment analysis and the predictions reached and making them more understandable. Furthermore, to achieve the vision and goal of the research, scientific steps have been implemented on several data sets, the most prominent of which is the Covid-19 data set which has produced a vast amount of public sentiment views regarding the virus and vaccination so we have used it in this work as a case study. This paper attempts to solve this issue by utilizing machine learning methods and explainable artificial intelligence (XAI) techniques to create a sentiment analysis framework, that not only achieves high sentiment prediction accuracy but also offers clear and understandable explanations for the underlying factors influencing sentiment classifications. The goal of the paper is to close the gap between human interpretability and accurate sentiment analysis by incorporating XAI approaches into machine-learning models.

Keywords: Sentiment analysis; Data Analysis; Explainable Artificial Intelligence.

1. Introduction

Sentiment analysis, a subfield of natural language processing (NLP), or opinion mining is the computer study of people's beliefs, opinions, sentiments, evaluations, and attitudes about various entities, such as goods, services, organizations, people, issues, occasions, and attributes. It offers insightful analysis by automatically classifying and analyzing opinions expressed in textual data [1]. The field's inception and explosive expansion align with the emergence of social media platforms on the Internet, including discussion boards, review sites, Twitter, and other social networks, which offer users access to a multitude of data expressing viewpoints. Sentiment analysis presents a lot of difficulties for user-generated textual online content mining, particularly in Arabic sentiment analysis. This is because Arabic is a complex language with a wide range of dialects, intricate morphology, and rich meanings. Many studies have worked on Arabic sentiment analysis and tried to find a unified model for data processing[2, 3, 4]. There is a chance to use machine learning methods for sentiment analysis because large Arabic text datasets are available[5]. These datasets have been gathered from social media and online forums. Effective public health policies can be informed by an understanding of public attitudes and opinions, particularly in Arabic, a language with multiple dialects. In this study, we use machine learning techniques to examine the emotions conveyed in Arabic literature. Sentiment analysis models are useful in many applications, but our capacity to comprehend their predictions determines how reliable they are. In response, there has been a recent upsurge in interest in XAI methods. In the context of sentiment analysis, XAI provides a route to interpretability and transparency, particularly about the Arabic language about COVID-19 and vaccination uptake. It can be difficult to understand the decision-making process of complex machine learning models since they frequently function as "black boxes." This paper introduces XAI methods which are designed to shed light on these mysterious models, XAI fills up the gaps between human comprehension and model performance by offering comprehensible explanations. Sentiment analysis with XAI enables us to determine the reasons for the emergence of a specific sentiment (positive, negative, or neutral) in Arabic text about any topic, particularly in the dataset related to COVID-19. The main contributions of our paper are as follows:

- Preprocessing Arabic text: Preprocessing the Arabic text is the first thing we do, including: Tokenization is the process of dividing a text into discrete words or units. Break word Removal is the process of getting rid of terms that are used frequently but don't have any meaning, such as "and" and "the". Stemming is the process of breaking down words to their most basic form for example, أبحاث → بحث . These procedures streamline the data and improve the analysis that comes after.
- Sentiment classification: This paper concentrates on categorizing Arabic text into three types of sentiments: neutral, negative, and positive. The machine learning methods we have selected are as follows: A supervised learning technique called Support Vector Machines (SVM) determines the ideal hyperplane to divide sentiment classes. Decision-based algorithms that divide data according to pertinent features are called Random Forests (RF) and Decision Trees (DT). A probabilistic classification model connecting sentiment is called logistic regression (LR). These algorithms have a successful track record in text classification tasks. Our work evaluates each method's accuracy, precision, and recall.
- The need for transparency: Accurate forecasts alone will not suffice to create trustworthy sentiment analysis; we also need to understand the logic underlying such predictions using XIA methods.

Here is how the rest of the paper is structured: Section 2 provides a brief background about the knowledge and technology used; Section 3 summarizes recent related works; Section 4 outlines our proposed approach methods; Section 5 introduces experimental findings; and Section 6 wraps up the paper by discussing possible future directions.

2. Background

This section introduces the main points of knowledge and technology that have helped us in our research, which includes sentiment analysis, sentiment techniques, Data augmentation, and Machine learning Models.

2.1 Sentiment analysis

A computational approach within the field of (NLP) aids in discerning the polarity of a given textual material, classifying it, for example, as positive, negative, or neutral[6]. The process entails the examination of perspectives expressed in written text to comprehend and assess corresponding responses. Sentiment analysis, alternatively referred to as opinion mining or emotion AI, is a computational technique used to analyze and interpret the sentiment or emotional tone expressed in textual data. Artificial intelligence encompasses a potent methodology that holds significant relevance in the realm of business, making it a crucial technology. It is also known as a valuable tool for monitoring social media platforms to gain insights into public sentiment on a specific product or service. Additionally, it can be utilized to analyze client comments and reviews to enhance the quality of products and services. Affective computing refers to the systematic application of natural language processing, text analysis, computational linguistics, and biometrics to identify, extract, measure, and analyze affective states and subjective information. As an illustration, in contemporary times, individuals seeking to make a consumer purchase are no longer constrained to seeking input solely from personal acquaintances. This is due to the proliferation of user evaluations and debates about the product, readily available in public online forums. In recent times, there has been a noticeable phenomenon where expressive content shared on social media platforms has played a significant role in influencing the structure of enterprises as well as shaping public opinions and emotions. These developments have had a profound impact on our social and political institutions.

2.2 Sentiment techniques

A set of methodologies is utilized to analyze the emotional and opinionated aspects present within textual content. SA uses different tools such as the Natural Language Toolkit (NLTK), Text Blob, Varder Sentiment, Google Cloud Natural Language API, Spy, Monkey Learn, Microsoft Azure Text Analytics API, and Amazon Comprehend. Several commonly used emotion analysis. techniques include: • Lexicon-based strategies involve the utilization of a lexicon or a corpus containing words that possess pre-established sentiment scores or polarity. As an illustration, the emotion of happiness can potentially be assigned as a positive numerical value, whereas the emotion of sadness can potentially be assigned as a negative numerical value. The sentiment of a given text is determined by evaluating the scores assigned to the individual words included within it. • Machine learning techniques are employed in sentiment analysis to leverage supervised or unsupervised learning algorithms. These algorithms are trained on labeled or unlabeled data, enabling them to discern the sentiment of a given text. For instance, neural networks, (SVM), and naive Bayes classifiers are among the algorithms commonly used in machine learning. • Hybrid techniques refer to the integration of lexicon-based and machine learning-based approaches, aiming to capitalize on the respective strengths of each method. Two hybrid systems, namely pSenti and SAIL, employ a combination of lexicons and neural networks to conduct sentiment analysis

2.3 Data augmentation

A methodology employed to expand the training set by generating modified replicas of a dataset through the utilization of existing data[7]. This process encompasses the implementation of slight modifications to the dataset or the utilization of deep learning techniques to generate novel data instances. Data augmentation techniques such as: • Word or Sentence shuffling: involves randomly changing the position of a word or sentence. • Word replacement: replaces words with synonyms. • Syntax-tree manipulation: paraphrase the sentence using the same word. • Random word insertion: insert words at random. • Random word deletion: delete words at random.

2.4 Machine learning

It is the systematic exploration and analysis of extensive sets of unprocessed data to identify recurring patterns and extract valuable insights. The utilization of machine learning techniques algorithms a range of strategies to transform extensive datasets into valuable outcomes. The most prevalent machine learning approaches encompass those used in this work:

SUPPORT VECTOR MACHINE (SVM): A supervised learning algorithm used for regression and classification applications is SVM. Its goal is to identify the best hyperplane for classifying the data. By utilizing various kernel types, SVM can solve classification issues that are either linear or non-linear. SVM can be computationally expensive for large datasets, but it is considered robust against overfitting and efficient at processing high-dimensional data. Depending on the particular issue and requirements, you can use a variety of performance indicators to assess the accuracy of (SVM) classifier. Classification accuracy is a frequently used metric that counts the percentage of cases that are correctly classified out of all instances. This study uses the Radial Basis Function (RBF) Kernel, also known as the Gaussian Kernel. It creates normal curves (bell-shaped functions) around data points. Summing these curves defines the decision boundary that allows complex decision boundaries from a mathematical perspective,

$$K(X, Y) = \tanh(\gamma \cdot X^T Y + r)$$

LOGISTIC REGRESSION (LR): A statistical model that is used for binary classification tasks. It estimates the probability of an instance belonging to a particular class using a logistic function and can handle linear and non-linear classification problems. It is widely used due to its simplicity, interpretability, and efficiency. Logistic regression assumes a linear relationship between the features and the log odds of the target variable.

DECISION TREE (DT): A hierarchical tree-like model is used for classification and regression tasks. It partitions the feature space based on the values of different features to make predictions. (DT) is easy to interpret and visualize, making it useful for understanding the decision-making process. It can handle both numerical and categorical features and can capture non-linear relationships. (DT) is prone to overfitting, especially when the tree depth is too large.

RANDOM FOREST (RF): It is defined as an ensemble learning method that combines multiple (DTs). It creates an ensemble of trees by training each tree on a random subset of the data and features. Random Forest improves generalization and reduces overfitting compared to a single (DT). It provides feature importance measures based on the average impurity reduction across all trees. (RF) is computationally efficient and can handle high-dimensional data.

3. Related work

Understanding public opinion and sentiment from textual data is made possible in large part by sentiment analysis. The rise in user-generated material on social media platforms, online reviews, and other sources, particularly during the COVID-19 pandemic, which has led to an expansion in the function of sentiment analysis in today's world. Many studies analyze English text; however, there is not much interest in sentiment analysis for the Arabic language. The most recent studies on sentiment analysis are explained in the following paragraphs. Alyami et al.[8] have built a sentiment analysis system with Twitter data. The SVM model is used in this work to categorize opinions in Arabic micro-texts as either positive or negative. When employing the SVM model, their experimental results yield an accuracy of 89.83%. Alshehri et al.[9] have examined how various models, such as MBERT, ARAELECTRA, MARBERT, ARBERT, Giga BERT, and others, affect analysis in Arabic text. The ARBERT model, a transformer-based model for Arabic language understanding, yields accuracy results of 92.39%, but it has a drawback in that it necessitates intricate manual feature engineering. Singh et al.[10] have used machine learning techniques, specifically convolutional neural networks (CNN) and long short-term memory networks (LSTM), to analyze Arabic sentiment classification on Twitter. The accuracy results are 89%. Basabain et al.[11] have created a program that uses the 83% accurate MARBERT algorithm to assess the tone of tweets about Saudi Arabian travel. Abd-Elaa et al.[12] have created a method with an 89% accuracy rate that utilizes (SVM) to identify ISIS online communities on Twitter. Tarek Gharib et al.[13] have achieved 53.83% accuracy ratings using LSTM-BILSTM models in an ensemble deep-learning method for Arabic tweet emotion identification. Lubna Abdelkareim et al.[14] have utilized CNN-LSTM to create a unique stacking ensemble of hybrid and deep learning models for enhanced Arabic sentiment analysis, achieving 81.4% accuracy on the Arabic sentiment tweets dataset (ASTD). Sherief Abdallah et al.[15] have enhanced sentiment analysis of the Emirati dialect. The results demonstrate that the best accuracy result is 80.80% when the ensemble model is used for the sentiment classification of the unbalanced dataset. Reem ALBayari et al.[16] have used (SVM) with a benchmark dataset based on Instagram to identify cyberbullying in Arabic, yielding results with a 69% accuracy rate. Martin Reisslein et al.[17] have examined COVID-19 tweets using a hybrid framework; the system is assessed using several metrics, including F1 score, recall, accuracy, and precision. According to our findings, the majority of people have either positive (38.5%) or neutral (34.7%) opinions. Moreover, the framework has chosen the (LSTM) neural network as its preferred machine learning technique, with an accuracy of 83%. Ruba Alhejaili et al.[18] have utilized many traditional machine learning techniques, including (DT), (RF), (SVM), and (LR). The best result is shown by (LR), which has an accuracy rate of 87%.

Table I represents comparison between various sentiment analysis methods employed in social media analysis. It highlights the research papers, techniques applied, performance metrics. It serves the objective of summarizing several research publications that deal with Arabic sentiment analysis that is run by machine learning or deep learning. The table contains the accuracy results and the year of publication. This table seeks to give a succinct summary of several methods and how well they work for Arabic sentiment analysis. It draws attention to the advantages and disadvantages of each study, including the application of certain models, the quantity and variety of datasets, the inclusion of error or qualitative analysis, and the attention given to linguistic complexity such as dialectal variations, sarcasm, or irony. Researchers and practitioners in the field of sentiment analysis can more easily compare studies and comprehend their contributions, limits, and possible areas for improvement thanks to the table's organized presentation of the data.

Table 1. Comparative Evaluation of Sentiment Analysis Approaches in Social Media

No	Authors	Methods	Results (Accuracy)	Year	Advantages	Disadvantages
1	Alyami et al[8]	(SVM)	89.83%	2020	Overcomes the limitations of current Arabic sentiment analysis methods, such as lack of resources.	It does not evaluate its performance against other cutting-edge techniques.
2	Alshehri et al.[9]	ARBERT	92.93%	2023	It improves the performance of sentiment and emotion classification on seven Arabic datasets, especially for small-resource settings.	It does not carry out any qualitative or error analysis to identify the advantages and disadvantages of the suggested models.
3	Singh et al.[10]	DT	88.51%	2021	It uses a large and diverse dataset of more than 14000 tweets collected from Kagle.	It does not take into account how emoticons, intensifiers, or negation affect the tweets' sentiment polarity.
4	Basabain et.a[11]	MARBERT	83%	2023	It addresses the need for sentiment analysis resources in the context of Saudi Arabia's tourism.	It uses a small dataset that consists of 2293 tweets.
5	A. Abd-Elaa et al[12]	SVM	89%	2020	The paper proposes an intelligent system that autonomously detects the ISIS online community on Twitter social media platforms. The paper analyzes both linguistic features and behavioral features such as hashtags, mentions, and who they follow.	The ethical and legal ramifications of identifying and blocking violent radical Twitter accounts are not discussed in the paper.
6	Tarek Gharib[13]	LSTM-BiLSTM	53.82%	2022	The paper combines three state-of-the-art deep learning models: Bi-LSTM, Bi-GRU, and MARBERT, which are based on recurrent neural networks and transformers.	The architecture, hyperparameters, optimization technique, and other aspects of the ensemble model's implementation and training are not well covered in the work.
7	Lubna Abdelkareim[14] et al	CNN-LSTM	81.4%	2022	The suggested model is assessed using three Arabic sentiment datasets (ASTD, Main-AHS, and Sub-AHS) and contrasted with a number of machine learning and deep learning models in this study.	The implementation and training of the stacking ensemble model, including the architecture, hyperparameters, optimization technique, and so on, are not sufficiently covered in the work.
8	Sherief Abdallah et al[15]	Classification models	80.8%	2022	To provide a useful resource for Arabic natural language processing, the study creates the first manually annotated dataset of the Emirati dialect on the Instagram platform.	A thorough explanation of the annotation methodology, sampling strategy, inter-annotator agreement, and other aspects of the comment collection and annotation process is lacked in this study.
9	Reem ALBayari et al[16]	SVM	69%	2022	The paper presents the first Instagram Arabic corpus for cyberbullying detection, which is a valuable resource for natural language processing and social media analysis.	The performance of the different learning models is not compared in the paper to baselines or other methods already in use for cyberbullying detection, such as lexicon-based or deep-learning models.
10	Martin Reisslein et al[17]	LSTM	83%	2022	The framework extracts and labels Twitter sentiments using the VADER lexicon-based method, which works well for brief and informal text data.	Topic modeling and infodemiological analysis, which can offer further insights into public views and behaviors during the pandemic, are not carried out by the framework of the tweets.
11	Ruba Alhejaili et al[18]	LR	87%	2021	Proper dataset collection and preprocessing are crucial for accurate sentiment analysis, and the paper provides details on these steps.	This study does not explore the generalizability of the trained models to other datasets or domains.

4. Proposed model

The objective of this paper is to propose an enhanced model for sentiment analysis specifically designed for Arabic text, utilizing machine learning techniques. The proposed model aims to overcome the linguistic intricacies of the Arabic language and enhance the accuracy and effectiveness of sentiment classification for Arabic text, also clarify the results, and make them more interpretable. Key Components of the Proposed Model.

4.1 Data collection

Collecting information from social media sites like (Twitter / X) that is pertinent to datasets that have been identified. Therefore, the primary reason we use the Twitter API is because of its policy, which formerly allowed users to use it but is currently unavailable. The Tweets vaccine data set, was gathered in February and March 2021 using terms associated with COVID-19 and people's opinions about getting the vaccine, particularly from Arabic countries such as Egypt, Saudi Arabia, and the United Arab Emirates.

4.2 Data preprocessing

Before conducting sentiment analysis on Arabic text, we need to preprocess the data to ensure that it is in a format suitable for analysis. The following preprocessing steps are recommended:

- **Normalization**: Normalize the Arabic text by converting all characters to their Unicode representation. This step ensures consistency across the dataset and helps mitigate issues related to different character encodings.
- **Diacritics Removal**: Remove any diacritical marks (e.g., vowel signs, gemination marks) from the text. These diacritics are not always consistently used in Arabic text and can introduce unnecessary noise into the sentiment analysis.
- **Stop Words Removal**: Identify and remove common stop words, such as conjunctions, prepositions, and articles, that do not contribute significantly to the sentiment of the text, and removing any language other than Arabic, as well as deleting all numbers, punctuation marks, and any other symbols such as (@, $, %, *, and #).
- **Tokenization**: Split the text into individual tokens (words) using an appropriate tokenization method for the Arabic language. This often involves handling word boundaries and dealing with attached particles and affixes.
- **Lemmatization**: To lower the dimensionality of the feature space and enhance the sentiment analysis model's generalizability, reduce each word to its root or dictionary form (lemma).
- **Negation Handling**: Recognize and deal with negation words and phrases, as they have a big effect on the text's tone.
- **Text Cleaning**: Take off all extraneous or distracting textual elements, like hashtags, email addresses, URLs, and special characters.
- **Stemming**: This process is launched immediately upon completion of the previous steps, which means returning the word to its roots. The goal is to clarify the original meanings of the words and their synonyms.

4.3 Sentiment analysis with text blob

One of the most important tasks in identifying and categorizing the sentiment expressed in textual data is sentiment analysis. This work covers two methods for sentiment analysis: one that makes use of the text Blob library and another that makes use of a machine learning model. The following steps are involved in using text Blob for sentiment analysis:

- **Text Blob Initialization**: initializing a text Blob object for each text instance by passing the preprocessed Arabic text data to the text Blob library.
- **Sentiment Polarity Detection**: determining the sentiment polarity of a text by using the sentiment property that the Text Blob library offers. Two values are used to express the polarity of sentiment: subjectivity, which runs from 0 (objective) to 1 (subjective), and polarity, which goes from -1 (negative sentiment) to 1 (positive sentiment).
- **Sentiment Classification**: categorizing a sentiment as positive, negative, or neutral based on its polarity value. By using text blob which gives polarity and subjectivity for new column (cleaned tweet) to determine sentiment analysis the result is not good which provides overfitting data so the percentage of neutral is near to 90% and positive and negative near to 10% from the total results.

4.4 Sentiment analysis with machine learning

Using machine learning approaches for sentiment analysis on the preprocessed Arabic text data to get over the drawbacks of language-specific models. The steps involved in this approach are as follows:

Feature extraction: is the procedure of choosing and transforming raw data into a set of pertinent features that effectively represent and describe the data, which is used to describe the process of turning textual data (sentences or texts) into numerical feature vectors that machine learning algorithms can comprehend. The goal of this stage is to extract pertinent textual information that can be utilized to categorize the input's sentiment and utilize feature extraction methods to represent the text with hand-crafted features, such as Bag-Of-Words (BOW). This may entail the extraction of syntactic data, statistical features, or domain-specific expertise.

Word embedding: is a method that uses automation to represent words in continuous space as low-dimensional, dense vectors. Word embeddings use the distributional features of words in a large corpus to capture the syntactic and semantic links between words and map the semantic ties between related words by mapping the word into an n-dimensional vector for the word's embedding to accurately convey its meaning. As a result, even if a model has only been trained on one word, it can still communicate with words that are similar to it. As a result, word embedding-based systems have shown great success in a range of natural language processing applications. It has many techniques such as (Word2vector– GloVe- FastText), It is discovered that FastText performs better than other methods when it comes to word embedding in Arabic (SA). In the current study, words are vectorized or transformed into machine-readable code using the FastText word embedding technique which learns both word embeddings and subword embeddings, representing words as a set of vectors. This enables it to process words that are not in its lexicon and to record morphological data. Recently, word vectors for 157 languages—including Arabic—have been pre-trained for FastText. It is a helpful tool for languages like Arabic that have intricate morphological structures since it takes into account the internal structure of words. Additionally, it offers high-quality embedding for uncommon or non-training words.

4.5 Data augmentation

One method that is frequently used to deal with the problems caused by a lack of labeled data in sentiment analysis assignments is data augmentation through increasing the training data and strengthening the generalization and robustness of our sentiment analysis models by creating more synthetic cases from the current dataset. The main reason for using data augmentation in this study is to make data used more balancing[19] by:

Minimizing Overfitting: The model's ability to adapt adequately to fresh, untested data is not always ensured by large datasets. So, by exposing the model to a wider range of examples during training, data augmentation can assist avoid overfitting and improve generalization.

Increasing Robustness: By adding more data, the model can withstand noise, disturbances, and other types of distribution shifts that could happen in real-world situations. The model can become more adept at handling these difficulties by being exposed to a greater variety of variations throughout training. Here, the data augmentation procedure is used following the sentiment analysis with the text Blob package. The following steps are included in the data augmentation process:

Original Dataset: Using the text Blob package, to perform sentiment analysis on the original labeled dataset.

Selection of Augmentation strategies: Choosing one or more augmentation strategies according to the features of the Arabic text data and the particular requirements of our sentiment analysis work so this research works with two strategies (Paraphrasing" rewriting sentences while preserving their meaning"- Text generation" create new synthetic instances").

Implementation of Augmentation: By using the chosen augmentation technique create new synthetic examples for every sentence in the original dataset. As a result, the dataset is enlarged and filled with more occurrences.

Combining and Equalizing: combining the enhanced dataset with the initial dataset to guarantee an equitable portrayal of sentiment classifications. By doing this action, bias towards a specific sentiment category is avoided and class distribution is maintained.

Testing and Validation Data: By accepting machine learning techniques and dividing the enhanced dataset into subsets for training, validation, and testing. After that, the sentiment analysis models are retrained and improved using the enhanced training set. Finally, we should determine that in the field of machine learning, data augmentation is frequently used with non-image data, such as text or tabular data, but in deep learning, data augmentation is commonly used for image data. That involves applying various transformations to the images to create new training examples.

4.6 Training / Testing

The training process involves several steps. First, the data collection step is responsible for gathering data and presenting it as a dataset. This data is then preprocessed through various cleaning and tokenization steps, such as removing Arabic stop words, removing punctuations, and performing stemming. After the data preparation, the training models step completes its stage by making the system use four different machine learning models (SVM, RF, LR, and DT). These models are trained on the prepared dataset.

Testing: The testing phase is represented by a separate box in the diagram. Here, the trained models are evaluated using various metrics like accuracy, recall, and precision. This testing stage allows the researchers to assess the performance of the different models and select the one(s) that best fit their requirements.

4.7 Explainable artificial intelligence (XAI)

Any machine learning model, regardless of complexity, can have its predictions interpreted using model-agnostic explanation ability techniques, that can utilize model-specific explain ability strategies which are customized to the machine learning algorithms used in our sentiment analysis models in addition to model-agnostic approaches. This paper determines using the following explainability techniques for the particular algorithms used (LR, SVM, DT, RF).

(LR): Look at the coefficients that are linked to every feature in logistic regression. Words or phrases linked to positive sentiment are indicated by positive coefficients, whereas those linked to negative sentiment are shown by negative coefficients.

(SVM): It is Algorithms that give judgment boundaries between several sentiment classes. The instances that are closest to the decision border are known as support vectors and show these boundaries. Comprehending the support vectors facilitates our understanding of the pivotal moments that ascertain the sentiment classification.

(DT): it is used as an interpretability to determine which features are the most discriminative and which decision rules are employed for sentiment classification, examine the (DT) structure. Understanding the decision-making process is facilitated by the decision tree's visual representation, which shows the path from the root to a particular emotion forecast.

(RF): Several decision trees make up Random Forest models, to determine which features are the most discriminative and which decision rules are employed for sentiment classification, by examining the decision tree structure. Understanding the decision-making process is facilitated by the decision tree's visual representation, which shows the path from the root to a particular emotion forecast that illustrates interpretability.

Algorithm # 1, the LIME Explainer Object
Input ⟵ Black-box mode, Sample instance
Output ⟵ Sample instance of a return explanation.
1. Begin
2. For given sample instance do:
3. Using a sample from the training dataset, a new dataset based on observation is created.
4. To find the difference between a permutation and the original data, use a formula.
5. Utilize a black-box model to predict the probability of novel points
6. Choose the m attributes that most accurately characterize the outcome of the complex model.
7. Utilize similarity as a linear mode weight for m-dimensional data.
8. In linear models, a choice's explanation is determined by its weights.
9. End For
10. End

Figure 1. Alogorithm for LIME

To gain deeper insights into the model's behavior, one can employ techniques like Local Interpretable Model-agnostic Explanations (LIME)[20] which is a popular technique that explains individual predictions made by a model. It works as follows:

- **Input sample**: To understand the model's prediction, start with a particular example of input (such as a product review or tweets), this study has used the tweets vaccine as a case study we take instances from its dataset.
- **Perturbation**: An input sample is disturbed or significantly modified, by LIME, which produces

a set of those variations. A portion of the input features, such as the text's words, are randomly masked off or otherwise altered to create these disturbances.

- **Model Prediction**: LIME derives the prediction from the machine learning model described for every perturbed input sample such as the indicated example on the positive instance, based on the probabilities which make a comparison between results from perturbed and original text.
- **Explanation Generation**: Based on the altered inputs and the accompanying model predictions, LIME then fits a straightforward, comprehensible model (such as a decision tree or linear regression) to the data. The forecast of the original model has an "explanation" in the form of this interpretable model.
- **Visualization**: LIME highlights the key elements (words) that helped the model anticipate the original input sample and display the explanation in a format that is easy for humans to understand. The pseudocode of LIME is explained in the figure 1.

Figure 2 depicts the XAI module for the standard ML technique, which can produce more transparent and reliable modeling [20].

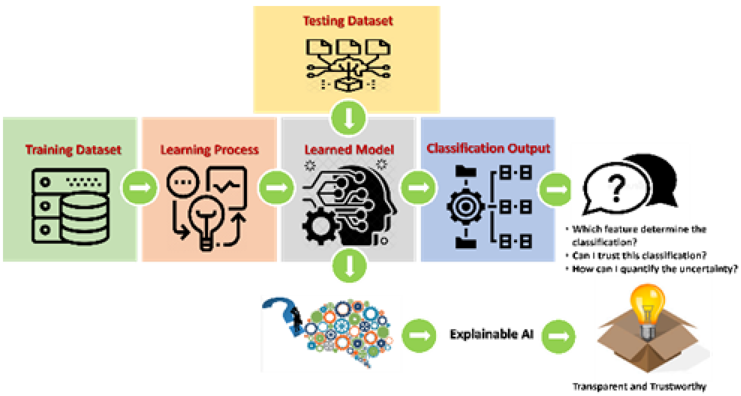

Figure 2. The overall model for XIA

Figure 3 represents the proposed model of our study that determines all steps that start with collecting data represented in the data set then preprocessing this data and after this executing feature extraction followed by a word embedding process (using fast text) and then training the machine learning models. After previous steps applying XIA (LIME) model to explain predications. Lastly, evaluate the performance machine learning models to determine (accuracy, recall, precision). (All of the procedures mentioned in this part, are applied sequentially, rather than concurrently to the data set.

5. Experimental results

5.1 Dataset Description

This section presents the experimental results obtained from the Tweets vaccine dataset besides two other datasets (the "100 K reviews" dataset was gathered in June and July 2016 from a large number of individuals who posted reviews in Arabic on the booking.com website using keywords associated with hotel reviews, "SS2030" dataset collected user opinions from Saudi Arabian Twitter users using keywords related to any development of Saudi Arabia in many sectors, this data was obtained over two months from October to November 2018) which used to prove that our model

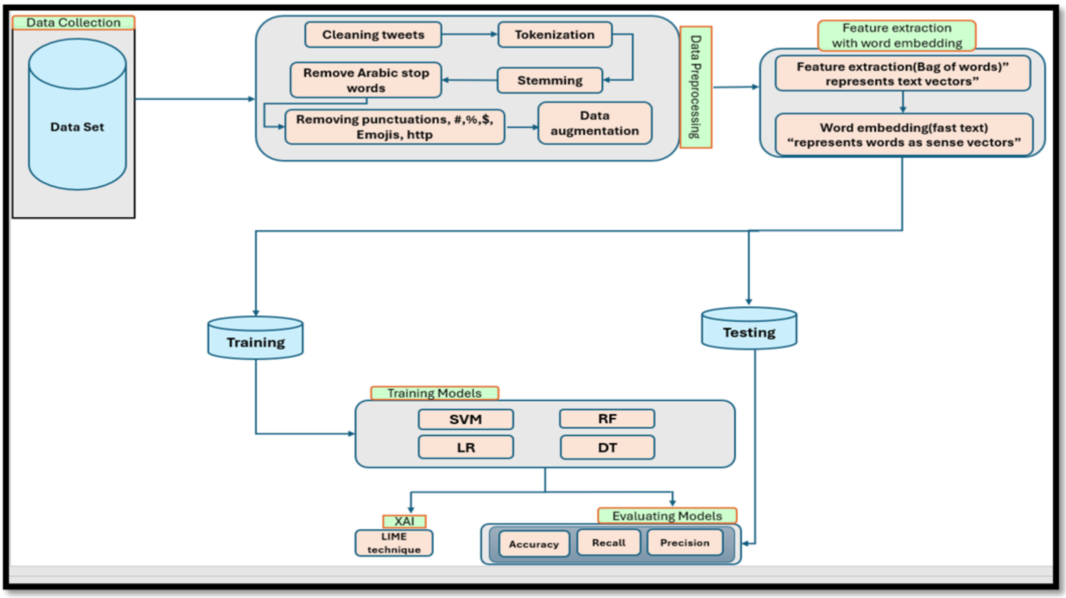

Figure 3. The proposed model

can work with different datasets and get the same results. Three distinct datasets (Tweets vaccine, 100K reviews, and SS2030) have been used to ensure transparency in the research process, we have compared the Tweets vaccine dataset with the other two datasets to show the presented results. **Tweets vaccine dataset** [21] : This dataset encompasses 32,476 Arabic tweets specifically related to COVID-19 vaccinations. It includes various attributes such as tweet date, handle, text, profile URL, name, and tweet link. Through sentiment analysis, we have analyzed the sentiment expressed in these tweets. **100K reviews dataset** [22]: This dataset comprises a collection of 99,999 Arabic tweets. Our sentiment analysis focused on extracting the sentiment expressed in these tweets. The dataset primarily consists of two columns: the tweet text and the associated sentiment label. **SS2030 dataset** [23]: The SS2030 dataset consists of reviews and comments reflecting people's opinions on the actions taken by the King of Saudi Arabia toward achieving their 2030 goals. It contains a total of 4,252 items arranged into two columns: sentiment and text. The sentiment column represents the sentiment expressed in the corresponding text. For a more detailed description, Table 2 provides a brief description of this dataset.

Table 2. Dataset description

#	Dataset name	Features	Size	Language	Reference
1	Tweets vaccine	tweet date, handle, text, profile URL, name, and tweet link	32,476 records	Arabic	[21]
2	100K reviews	Text	99,999 records	Arabic	[22]
3	SS2030	Text	4,252 records	Arabic	[23]

5.2 Data Preparation

The first step involves preparing the three datasets (Tweets vaccine, 100K reviews, SS2030) for analysis. By running tasks such as data cleaning, preprocessing, and formatting to ensure the data is in a suitable format for the subsequent analysis.

5.3 Algorithm selection

For each dataset, four distinct algorithms are selected to carry out the sentiment analysis. The following methods (LR, SVM, DT, and RF) are employed, these algorithms have been probably chosen because they have worked well on sentiment analysis tasks and fit well with the properties of the data.

5.4 Model training and evaluation

A suitable training technique, such as supervised learning (a subset of machine learning that employs labeled datasets to train algorithms to predict outcomes), is applied to each algorithm to train it on the specific dataset. For the algorithms to learn the relationships and patterns between the sentiment labels and the text data, labeled data has to be fed to them during the training phase. Assessment criteria including accuracy, recall, and precision are used to assess the models following training.

5.5 Performance Metrics

The full experiments are carried out with a Python-based machine-learning framework. By dividing the dataset into training, validation, and testing sets at random with a ratio of 70:15:15, respectively, to guarantee accurate and objective findings. By using the following common evaluation metrics determined by the following abbreviations ("TP" True Positive, "TN" True Negative, "FP" False Positive, "FN" False Negative) to assess the sentiment analysis model performance, the following results will be attained: Accuracy: The percentage of tweets in the testing set that are properly classified out of the total number of tweets.

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

Precision: defined as the proportion of accurately predicted positive tweets to all tweets that are projected to be positive

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

Recall: The percentage of positive tweets that are accurately predicted to all positive tweets that are sent out.

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

The confusion matrix is a basic metric used in our research to assess how well our classification models performed on the Arabic tweet's vaccination dataset. A thorough overview of the model's predictions and the actual labels assigned to the data is given by the confusion matrix. It enables us to compute a range of assessment measures, including recall, accuracy, and precision. It is also known as a square matrix. Concerning the model's predictions and the ground truth labels, each cell indicates the number or percentage of cases that fall into a specific category of the cells determined in Figure 4.

5.6 Results summarizations

This study presents the experimental results in a series of tables for each of the three datasets used (Tweets vaccine dataset, 100K reviews dataset, SS2030 dataset). Table 3 summarizes the results for

		ACTUAL VALUES	
PREDICATED VALUES		POSITIVE	NEGATIVE
	POSITIVE	TP	FP
	NEGATIVE	FN	TN

Figure 4. Confusion matrix design

the Tweets vaccine dataset. It shows the accuracy, recall, and precision scores for the four machine learning algorithms evaluated (LR, SVM, DT, RF). The SVM algorithm achieves the highest accuracy of 90.7%, followed by LR at 89.7%, RF at 89.5%, and DT at 79%.

Table 3. EXPERIMENTAL RESULTS OF THE FIRST DATASET (TWEETS VACCINE)

Algorithm/Results	LR	SVM	DT	RF
ACCURACY	89.7%	90.7%	79%	89.5%
RECALL	88.5%	89%	77%	88%
PRECISSION	89.4%	90%	78%	89%

Table 4 displays the results for the 100K reviews dataset. Again, the performance metrics of accuracy, recall, and precision are reported for the four algorithms. LR obtains the highest accuracy of 90.9%, SVM is close behind at 90%, while DT and RF have lower accuracies of 77.8% and 85.5% respectively.

Table 4. EXPERIMENTAL RESULTS OF THE FIRST DATASET (100 K reviews)

Algorithm/Results	LR	SVM	DT	RF
ACCURACY	90.9%	90%	77.8%	85.5%
RECALL	86.3%	87%	76.5%	80%
PRECISSION	88%	86.7%	76.9%	83%

Finally, Table 5 presents the results for the SS2030 dataset. SVM emerges as the top performer with an accuracy of 91.1%, outpacing LR at 88.5%, RF at 86%, and DT at 83%.

Table 5. EXPERIMENTAL RESULTS OF THE FIRST DATASET (SS2030)

Algorithm/Results	LR	SVM	DT	RF
ACCURACY	88.5%	91.1%	83%	86%
RECALL	87%	89.5%	81.5%	84.8%
PRECISSION	88%	89%	82.3%	85%

Overall, the results indicate that the (SVM or LR) algorithms, when working separately with the data processing techniques employed in this study, achieve the best performance across the three datasets in terms of accuracy, recall, and precision.

5.7 Comparison with previous investigations

Our suggested model, which uses (SVM - LR - DT- RF) models has determined that the SVM and LR algorithms with the data processing processes we carry out throughout the research get the best results, as shown in the preceding figures and tables. When comparing these results with results in the previous studies in terms of performance on all three datasets, the proposed model is better than those used in previous studies. Table 6 provides a comparison of accuracy achieved by various sentiment analysis algorithms on different datasets (Tweets vaccine, 100K reviews, and SS2030) and our proposed model. Because this study focuses on Machine learning only, we have chosen the previous works that have focused only on machine learning. For the "Tweets vaccine" dataset, the previous algorithms considered are (LR) and Support Vector Classification (SVC), and our proposed model determined better results achieved by SVM and LR. The accuracies achieved are 87% for LR, 84% for SVC, and 90.7% and 89.7% for our proposed model (SVM, LR) respectively. Similarly, for the "100 K reviews" dataset, the algorithms evaluated are (DT, and SVM), and a proposed model determines better results achieved by LR and SVM. The accuracies obtained are 74% for DT, 82.5% for SVM, and 90.9% and 90% for the proposed model (LR, SVM) respectively. Lastly, for the "SS2030" dataset, the algorithms analyzed are SVM, and our proposed model determines better results achieved by SVM and LR. The accuracies achieved are 89.83% and 91.1%, 88.5% for our proposed model (SVM, LR) respectively. This table serves as a valuable resource for researchers and practitioners interested in understanding the performance of different sentiment analysis algorithms on the datasets that are used in our work.

Table 6. COMPARATIVE ANALYSIS OF SENTIMENT ANALYSIS ALGORITHMS ON DIFFERENT DATASETS

Dataset name	Previous Work	Accuracy	Proposed model	Accuracy
Tweets vaccine	LR[18]	87%	LR	89.7%
	Support Vector Classification[24]	84%	SVM	90.7%
100 K reviews	DT[25]	74%	LR	90.9%
	SVM[26]	82.5%	SVM	90%
SS2030	SVM[8]	89.83%	LR	88,5%
			SVM	91.1%

5.8 Explainable Artificial intelligence

This work is interested in utilizing (XAI) techniques to interpret and understand the results obtained from the machine learning models, particularly SVM and LR. The following steps illustrate how to use XAI techniques especially LIME model in analysis:

Interpretability of SVM and LR Models: SVM and LR are considered relatively interpretable models compared to other complex algorithms like deep neural networks. They provide insights into the importance of features and coefficients associated with each class. For a given instance, LIME generates a perturbed dataset by randomly sampling similar instances, it trains a simple interpretable model (LR) on this perturbed dataset. For carrying out our work, we have taken a random sample from the (Tweets vaccine) dataset which is used as a case study in our investigation. This sample is estimated at 6161 records, the results can be represented as follows:

A sentiment analysis visualization as the following: True predictions for random examples from tweets vaccine data set which represents (Positive - negative - neutral) instances are shown in Figure 5. Figures (6,7,8) have a text excerpt with highlighted words indicating their impact on the sentiment score, as well as a bar graph displaying the prediction probabilities for positive, neutral, and negative attitudes.

positive	negative	neutral
أنصح الجميع بأخذ لقاح كورونا	مخاوف حول الآثار الجانبية للقاح كورونا	توفير اللقاح من عدمه أمر غير مجدي

Figure 5. Random examples from tweets vaccine dataset

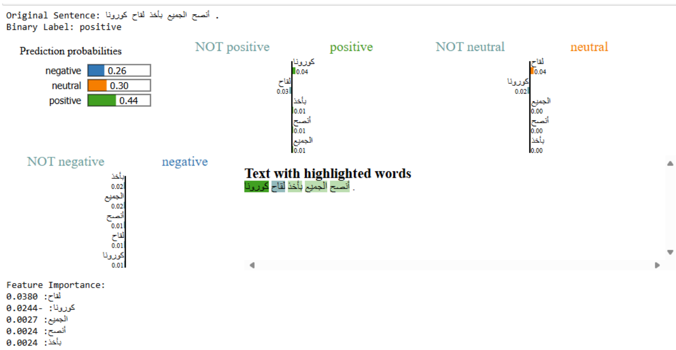

Figure 6. Sentiment Analysis Visualization (Predicted Class Probabilities for a Positive Sentence).

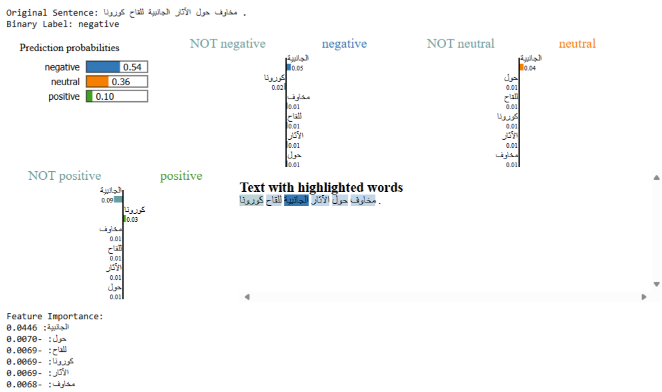

Figure 7. Sentiment Analysis Visualization (Predicted Class Probabilities for a Negative Sentence).

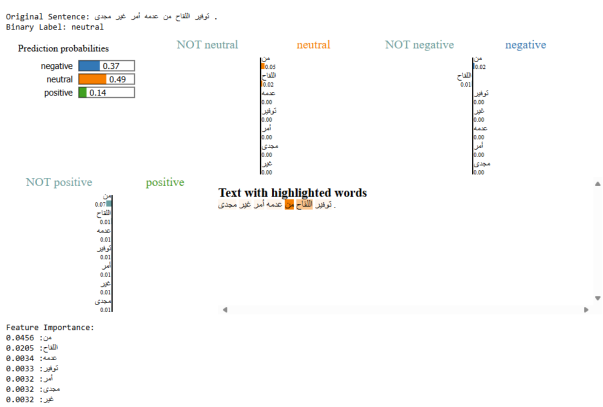

Figure 8. Sentiment Analysis Visualization (Predicted Class Probabilities for a Neutral Sentence).

False predictions for sentiment analysis are also represented by LIME in this work, for example Figure 9 shows a difference in the results that determined the true results are positive but the predicted is neutral.

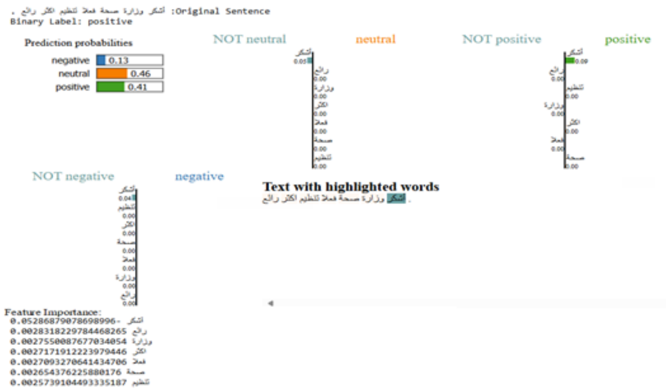

Figure 9. Sentiment Analysis Visualization (Sample of Difference between predicted and true results)

The coefficients of this interpretable model indicate the importance of different features for the specific prediction. By comparing the predictions of the original model with those of the interpretable model, LIME explains the specific instance.

The study's conclusions are derived from a comparison of 6161 records in the (Tweets vaccine) dataset by using the technique (partial dependence plot). This plot visualizes the relationship between specific input variables and the model's output Figure 10 presents a bar graph that compares the number of (positive, negative, neutral) for two types of labels: "Predicated labels" and "True labels". The graph shows that for the "Predicated labels", there are 1246 positive instances, 1611 negative instances, and 3302 neutral instances. For the "True labels", there are 1452 positive instances, 1660 negative instances, and 3039 neutral instances. The graph provides a visual representation of the distribution of the different label types, allowing for a clear comparison between the predicted and true labels. This type of information can be useful in evaluating the performance of a machine learning model or classifying data based on these label categories.

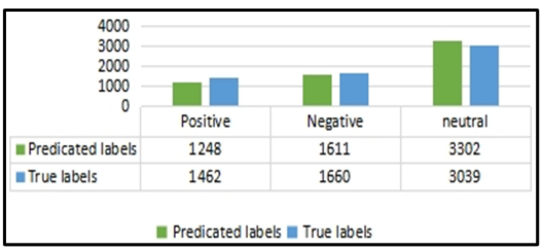

Figure 10. Distribution of Predicted and True Labels in the Tweets Vaccine Dataset Using Logistic Regression

True predictions represented in Figure 11 which is made by our model. The figure shows a stacked bar chart that compares "Predicted" and "True" labels for positive, negative, and neutral instances. The chart provides a visual representation of how the predicted labels align with the true labels across these three categories. The bars for "Predicted" and "True" are stacked on top of each other, allowing for a side-by-side comparison of the values. For the positive instances, the Predicted and True labels appear to be well-aligned. However, for the negative and neutral instances, there are some discrepancies between the Predicted and True labels, as indicated by the offsets between the stacked bars. The green line represents the expected labels, and the yellow dashed line represents the real labels. When the green line and the yellow dashed line closely match, the predictions match the true labels, demonstrating the model's strong performance. This type of visualization can be useful for evaluating the performance of a model or classification system, as it highlights areas where the predictions may be diverging from the ground truth labels. By understanding these differences, improvements can be made to enhance the accuracy and reliability of the predictions.

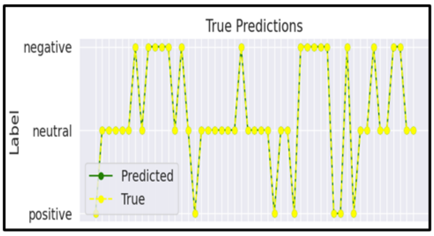

Figure 11. Analysis of True Predictions in the Tweets Vaccine Dataset Using Logistic Regression

False predictions are represented in Figure 12 which analyzes the "Predicted" and "True" labels for positive, negative, and neutral instances in a dataset. The focus of this visualization appears to be on cases where the predictions are incorrect, or "False Predictions." The stacked bars show the Predicted and True label values for each instance, with the Predicted label in red and the True label in yellow. Where the two colors do not align, it indicates a mismatch between the predicted and true labels - a false prediction. The chart highlights areas where the model's predictions diverge from the ground truth, which can be valuable information for evaluating and improving the model's performance. By identifying patterns in the false predictions, the model developers can work to address the weaknesses and enhance the overall accuracy of the system. This type of visualization provides a clear way to inspect and analyze the model's mistakes, which is an important step in the iterative process of refining and enhancing machine learning models.

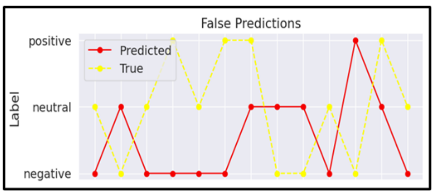

Figure 12. Analysis of False Predictions in the Tweets Vaccine Dataset Using Logistic Regression

In conclusion, XAI more especially LIME, is used in this investigation to clarify and analyze the findings. The utilization of machine learning methodologies, specifically concentrating on the LR model, produces good results. These results help users to make a great deal about how the machine learning models anticipate specific (tweets, product reviews, ...) by using LIME on the text datasets.

6. Conclusion and future work

This study has used machine learning methods, represented by (SVM-LR-DT-RF) to measure the accuracy of sentiment analysis for users of social media platforms and digital platforms, given their massive spread in all circles, specifically for Arabic-speaking users, who suffer in particular due to the lack of many types of research interested in analyzing Arabic texts due to their overlapping, difficulty, and the existence of many synonyms for it. Hence, this study is important not only for analyzing sentiment analysis of Arabic texts but also for improving the previous results reached and adding a feature that may enable new researchers to develop such results. Moreover, XAI is crucial to illustrate the final results issued by using machine learning models and to make them more understandable in a simplified and clear way. It is worth noting that although the work focused on the COVID-19 data set due to its noticeable impact on society, more than one data set has been used within our work to prove the success of our methods in dealing with different data sets with changing interests. Future studies can investigate sophisticated preprocessing methods designed exclusively for it. This can involve tackling dialectal variances, handling more complex stemming algorithms, and taking care of morphological issues. Subsequent studies may examine the integration of contextual data into sentiment analysis models, such as temporal factors, location, or user demographics. This can provide a more profound understanding of how attitudes change over time, vary by location, or change depending on demographic factors.

Open data statement

The Tweets vaccine dataset is available in the kaggle repository at https://www.kaggle.com/datasets/mahdimahdi55/data-tweet-s?resource=download, the 100K reviews dataset at https://www.kaggle.com/datasets/abedkhooli/arabic-100k-reviews, and the SS2030 dataset at https://www.kaggle.com/datasets/snalyami3/arabic-sentiment-analysis-dataset-ss2030-dataset.

References

[1] Souha Katat, Chamseddine Zaki, Hussein Hazimeh, Ibrahim El bitar, Rafael Angarita, and Lionel Trojman. “Natural Language Processing for Arabic Sentiment Analysis: A Systematic Literature Review”. In: *IEEE Transactions on Big Data* (2024), pp. 1–18.

[2] Ohud Alsemaree, Atm S Alam, Sukhpal Singh Gill, and Steve Uhlig. “Sentiment analysis of Arabic social media texts: A machine learning approach to deciphering customer perceptions”. In: *Heliyon* (2024).

[3] Yassir Matrane, Faouzia Benabbou, and Nawal Sael. “A systematic literature review of Arabic dialect sentiment analysis”. In: *Journal of King Saud University-Computer and Information Sciences* (2023), p. 101570.

[4] Arwa Alqarni and Atta Rahman. “Arabic Tweets-Based Sentiment Analysis to Investigate the Impact of COVID-19 in KSA: A Deep Learning Approach”. In: *Big Data and Cognitive Computing Vol. 7,* (2023), p. 16.

[5] Dhiaa A Musleh, Ibrahim Alkhwaja, Ali Alkhwaja, Mohammed Alghamdi, Hussam Abahussain, Faisal Alfawaz, Nasro Min-Allah, and Mamoun Masoud Abdulqader. “Arabic Sentiment Analysis of YouTube Comments: NLP-Based Machine Learning Approaches for Content Evaluation”. In: *Big Data and Cognitive Computing* 3 (2023), p. 127.

[6] Kian Long Tan, Chin Poo Lee, and Kian Ming Lim. “A survey of sentiment analysis: Approaches, datasets, and future research”. In: *Applied Sciences* (2023), p. 4550.

[7] Lucas Francisco Amaral Orosco Pellicer, Taynan Maier Ferreira, and Anna Helena Reali Costa. “Data augmentation techniques in natural language processing”. In: *Applied Soft Computing* (2023).

[8] Sarah N. Alyami and Sunday O. Olatunji. “Application of Support Vector Machine for Arabic Sentiment Classification Using Twitter-Based Dataset”. In: *Journal of Information and Knowledge Management* (2020), pp. 4–5.

[9] Wafa Alshehri, Nora Al-Twairesh, and Abdulrahman Alothaim. “Affect Analysis in Arabic Text: Further Pre-Training Language Models for Sentiment and Emotion”. In: *Applied Sciences (Switzerland)* (2023).

[10] Jyotsna Singh and Pradeep Tripathi. "Sentiment analysis of Twitter data by making use of SVM, Random Forest and Decision Tree algorithm". In: *Proceedings - 2021 IEEE 10th International Conference on Communication Systems and Network Technologies, CSNT 2021* (2021), pp. 193–198.

[11] Seham Basabain. "SAtour: Sentiment analysis of Saudi Arabia Tourism Tweets". In: (2023), pp. 11–12.

[12] Ahmed I.A. Abd-Elaal, Ahmed Z. Badr, and Hani M.K. Mahdi. "Detecting violent radical accounts on Twitter". In: *International Journal of Advanced Computer Science and Applications* (2020), pp. 516–522.

[13] Alaa Mansy, Sherine Rady, and Tarek Gharib. "An Ensemble Deep Learning Approach for Emotion Detection in Arabic Tweets". In: *International Journal of Advanced Computer Science and Applications* (2022), pp. 980–990.

[14] Hager Saleh, Sherif Mostafa, Lubna Abdelkareim Gabralla, Ahmad O. Aseeri, and Shaker El-Sappagh. "Enhanced Arabic Sentiment Analysis Using a Novel Stacking Ensemble of Hybrid and Deep Learning Models". In: *Applied Sciences (Switzerland)* (2022).

[15] Arwa A. Al Shamsi and Sherief Abdallah. "Sentiment Analysis of Emirati Dialects". In: *Big Data and Cognitive Computing* (2022).

[16] Reem Albayari and Sherief Abdallah. "Instagram?Based Benchmark Dataset for Cyberbullying Detection in Arabic Text". In: *Data* (2022).

[17] Piyush Vyas, Martin Reisslein, Bhaskar Prasad Rimal, Gitika Vyas, Ganga Prasad Basyal, and Prathamesh Muzumdar. "Automated Classification of Societal Sentiments on Twitter With Machine Learning". In: *IEEE Transactions on Technology and Society* (2021), pp. 100–110.

[18] Ruba Alhejaili, Enas S. Alhazmi, Abdullah Alsaeedi, and Wael M.S. Yafooz. "Sentiment Analysis of the Covid-19 Vaccine for Arabic Tweets Using Machine Learning". In: *2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions), ICRITO 2021* (2021), pp. 1–4.

[19] Kiran Maharana, Surajit Mondal, and Bhushankumar Nemade. "A review: Data pre-processing and data augmentation techniques". In: *Global Transitions Proceedings* (2022), pp. 91–99.

[20] Hanaa Salem, Ibrahim M El-Hasnony, A E Kabeel, Emad M S El-Said, and Omar M Elzeki. "Deep Learning model and Classification Explainability of Renewable energy-driven Membrane Desalination System using Evaporative Cooler". In: *Alexandria Engineering Journal* (2022), pp. 10007–10024.

[21] Mahdi ALmamlouk. *Tweeet Vaccine*. [Online; accessed March-2024]. 2021. URL: https://www.kaggle.com/datasets/mahdimahdi55/data-tweet-s/data.

[22] Abed Khooli. *Arabic 100k Reviews*. [Online; accessed March-2024]. 2020. URL: https://www.kaggle.com/datasets/abedkhooli/arabic-100k-reviews.

[23] Sarah Alyami. *Arabic Sentiment Analysis Dataset SS2030 Dataset*. [Online; accessed March-2024]. 2019. URL: https://www.kaggle.com/datasets/snalyami3/arabic-sentiment-analysis-dataset-ss2030-dataset.

[24] Lama Alsudias and Paul Rayson. "COVID-19 and Arabic Twitter: How can Arab world governments and Public Health Organizations learn from social media?" In: *Proceedings of the Annual Meeting of the Association for Computational Linguistics* (2020), pp. 23–25.

[25] Noor Khalid Bolbol and Ashraf Yunis Maghari. "Sentiment analysis of arabic tweets using supervised machine learning". In: *Proceedings - 2020 International Conference on Promising Electronic Technologies, ICPET 2020* (2020), pp. 89–93.

[26] Muhammad Mazen Almustafa2 Ali Mohammad Salloum1. "ANALYSIS AND CLASSIFICATION OF CUSTOMER REVIEWS IN ARABIC USING MACHINE LEARNING AND DEEP LEARNING". In: *Data Acquisition and processing* (2023), pp. 49–52.