Enhanced Emotion Analysis Model using Machine Learning in Saudi Dialect: COVID-19 Vaccination Case Study

Abdulrahman O. Mostafa, Tarig M. Ahmed
Department of Information Technology-Faculty of Computing and Information Technology
King Abdulaziz University, Jeddah 21589, Saudi Arabia

***Abstract*—Sentiment Analysis (SA) and Emotion Analysis (EA) are effective areas of research aimed to auto-detect and recognize the sentiment expressed in a text and identify the underpinning opinion towards a specific topic. Although they are often considered interchangeable terms, they have slight differences. The primary purpose of SA is to find the polarity expressed in a text by distinguishing between positive, negative, and neutral opinions. EA is concerned with detecting more emotion categories, such as happiness, anger, sadness, and fear. EA allows the analysis to extract more accurate and detailed results that suit the field in which it is applied. This work delves into EA within the Saudi Arabian dialect, focusing on sentiments related to COVID-19 vaccination campaigns. Our endeavor addresses the absence of research on developing an effective EA machine-learning model for Saudi dialect texts, particularly within the healthcare and vaccinations domain, exacerbated by the lack of an EA manual-labeled corpus. Using a systematic approach, a dataset of 33,373 tweets is collected, annotated, and preprocessed. Thirty-six machine learning experiments encompassing SVM, Logistic Regression, Decision Tree models, three stemming techniques, and four feature extraction methods enhance the understanding of public sentiment surrounding COVID-19 vaccination campaigns. Our Logistic Regression model achieved 74.95% accuracy. Findings reveal a predominantly positive sentiment, particularly happiness, among Saudi citizens. This research contributes valuable insights for healthcare communication, public sentiment monitoring, and decision-making while providing labeled-corpus and ML model comparison results for improving model performance and exploring broader linguistic and dialectal applications.**

Keywords—Data mining; natural language processing; sentiment analysis; emotion analysis; machine learning; support vector machine; logistic regression; decision tree; Covid-19

I. Introduction

Microblogging has grown significantly as a means of communication and information sharing, notably on platforms like Twitter. Offering real-time accessibility from anywhere, Twitter allows users, through short texts or "Tweets," to express thoughts, opinions, and emotions [1]. This social media giant has become a global hub for diverse purposes, including news updates, social interactions, and discussions. In Saudi Arabia, where Twitter boasts over 15.5 million active users [2], it is a valuable repository for researchers keen on comprehending public sentiments.

The advent of Natural Language Processing (NLP) techniques, such as Sentiment Analysis (SA) and Emotion Analysis (EA), has revolutionized the understanding of extensive textual data generated on platforms like Twitter. SA discerns a text's sentiment or emotional tone, while EA goes a step further to identify specific emotions like happiness, anger, or sadness. These techniques find widespread applications in diverse domains, including social media analytics [5], customer feedback analysis [6], and market research, offering insights into human opinions, emotions, and behaviors.

The research gap addressed in this study emerges from several significant factors within the context of emotion analysis in the Saudi dialect. Arabic is known for its rich morphology and many dialects [3]. Social media, particularly Twitter, introduces informal language, including dialects and slang, complicating the accurate interpretation of emotions [2]. Additionally, the absence of a multi-emotion class Saudi dialect labeled-tweets corpus, the presence of diacritical marks (Tashkeel) introducing ambiguity, and the deviation of the Saudi dialect from Modern Standard Arabic writing norms collectively contribute to this research gap. The study's overarching objective is to fill this void by crafting a machine-learning model adept at classifying Saudi dialect tweets into seven emotion categories, explicitly focusing on emotions related to COVID-19 vaccinations. Such an endeavor is poised to enhance our understanding of the sentiments expressed in Saudi Arabia concerning COVID-19 vaccinations.

Existing studies have provided limited insights using narrow classifications, making a more comprehensive range of classifications necessary for more valuable and effective results. Embracing EA offers a more comprehensive understanding of sentiments beyond binary or ternary classifications, particularly in contexts like COVID-19 vaccination campaigns.

The motivations behind the study are twofold. First, there is a need to focus on implementing an EA model in the Saudi dialect, anticipating its profound influence across various sectors such as business, healthcare, education, government, and technology. Second, to better understand the general attitudes towards COVID-19 vaccination campaigns held in Saudi Arabia. Objectives are aligned with these motivations, striving to produce an effective machine-learning model, enhance the accuracy of existing EA studies, unveil prevailing attitudes, create a Saudi dialect labeled-tweets corpus, and evaluate different algorithms comprehensively.

In anticipation of these objectives, the study expects to deliver a machine-learning model proficiently classifying Saudi dialect tweets into seven emotion categories. Additionally, it aims to create a labeled-tweets corpus, shedding light on the emotions expressed in diverse contexts, particularly in the healthcare and COVID-19 vaccination domain. Visual representations of general attitudes and detailed statistics are poised to empower decision-makers with nuanced insights, influencing policies and communication strategies. The model and corpus crafted in this study are envisioned to be valuable assets, not only for this research domain but also for broader applications in related studies.

Our paper is structured as follows: In Section II, we provide an overview of sentiment analysis and emotion analysis, highlighting their intersections and briefly discussing prior academic work in the field. Section III details the methodology, materials, and steps in constructing the final machine learning model, covering the dataset collection, annotation, and balancing methods. Section IV outlines the implementation of three machine learning models. Section V covers the evaluation methods and discusses the results achieved. Finally, in Section VI, we conclude the paper, presenting a summary and outlining our future work.

II. Background

A. Sentiment Analysis

1) Using machine learning: In health-related contributions addressing the COVID-19 pandemic, Aljameel et al. [18] developed a machine learning model gauging individuals' awareness of preventive measures during the quarantine in Saudi Arabia. Utilizing a dataset from the curfew period, they employed SVM, NB, and KNN classifiers, optimizing 85% accuracy by combining TF-IDF with SVM. Focused on the Saudi dialect, the study by Al Sari et al. [19] in the entertainment field used MLP, NB, SVM, RF, and Voting algorithms, achieving 90% accuracy with NB and MLP on Twitter data. Alhuri et al. [20] utilized GRU in an RNN, reaching an 81% F1 score for public reactions to COVID-19 in Arabic tweets. Alahmary, Al-Dossari, and Emam. The study in [4] outperformed ML with DL algorithms (Bi-LSTM) on the Saudi Dialect Twitter Corpus, achieving 94% accuracy. AlYami and AlZaidy [21] focused on Arabic Dialect Identification using SVM, RF, NB, and LR, achieving a maximum of 87% and 86% accuracy in Egyptian dialects using LR and SVM, respectively. However, they have worked on sentiment analysis fields only; limitations included a small dataset, the absence of manual annotation, and intensive preprocessing.

2) Using lexicon-based: Assiri, Emam, and Al-Dossari [3] proposed a domain-specific lexicon-based algorithm for the Saudi dialect, addressing the absence of such models. However, limitations arose from evaluating it against a non-Saudi dataset and focusing solely on text polarity. Similarly, Al-Thubaity et al. [10] manually created "SauDiSenti," a Saudi dialect sentiment lexicon, but challenges emerged in comparing it to a broader Arabic dictionary. Al-Ghaith [22] adopted a distinct approach, enhancing sentiment analysis accuracy by directly applying preprocessing tasks to the Saudi dialect lexicon, achieving 81% accuracy by relying on an English lexicon for the original creation.

3) Using hybrid approach: Very few published studies have utilized the Hybrid approach of SA in Arabic - where semantic orientation and ML techniques are combined. Aldayel and Azmi [9] used this approach to improve the F-measure score; they achieved an overall F-measure and accuracy of 84% and 84.01%, respectively. Alhumoud, Albuhairi, and Altuwaijri [23] used the same approach of combining two ML algorithms and applied the SA on 3000 Saudi dialect tweets to prove the efficiency of the hybrid learning approach compared to solo ML.

B. Emotion Analysis

EA can be described as recognizing distinct human emotions in contrast to Sentiment Analysis, which identifies whether data is positive, negative, or neutral [24]. Because of the lack of a labeled corpus for Saudi dialect that can be used for classifying emotions and polarity behind a text, Al-Thubaity, Alharbi, Alqahtani and Aljandal [10] introduced the Saudi Dialect Twitter Corpus (SDTC) that contains 5400 tweets of Saudi dialect and MSA classified for sentiment analysis and emotion analysis annotated by three raters based on their polarity (positive, negative, and neutral) for the sentiment, and based on Ekman's basic emotions (anger, fear, disgust, sadness, happiness, surprise, no emotion and not sure) for the emotion analysis. However, no ML or lexicon-based approaches have been applied to this corpus to evaluate its efficiency in this study. Another study by A. AlFutamani and H. Al-Baity [11] was the first in the EA in Arabic, focusing on Saudi dialects in Arabic textual content retrieved from Twitter, mainly in Saudi-based tweets. They built a system that can detect the underlying emotions of Saudi dialect tweets to classify them based on seven emotion categories (happiness, fear, disgust, anger, surprise, optimism, and sadness). They used two ML algorithms (SVM and MNB), achieving 73.39% accuracy in the SVM approach. However, they applied the analysis in dataset domain sets different from ours with varying models of ML.

In summarizing the related work, it is evident that sentiment analysis and emotion analysis have made significant strides, particularly in applications related to COVID-19 discussions on social media. However, within the context of our study, there exists a notable research gap. Previous works have primarily addressed sentiment analysis in broader contexts, lacking the depth to decipher the intricate emotional expressions specific to the Saudi population. Furthermore, the scarcity of labeled datasets in the Saudi dialect poses a considerable challenge. Our research seeks to bridge this gap by offering a comprehensive analysis of emotion analysis, utilizing a meticulously annotated dataset tailored to the Saudi dialect. This approach contributes to the broader field of sentiment analysis and provides a nuanced understanding of the emotional landscape surrounding COVID-19 vaccinations in Saudi Arabia.

C. Machine Learning Algorithms

ML algorithms encompass supervised, unsupervised, and reinforcement learning. Supervised learning trains on labeled data, associating input with output labels. It excels in classification, categorizing data into known classes, regression for predicting continuous values, and ranking for ordering data [7] [8]. Techniques like Support Vector Machines, Logistic Regression, Decision Trees, Random Forests, and Neural Networks are part of supervised learning, each chosen based on factors like data nature and problem complexity. Unsupervised learning handles unlabeled data, discerning patterns without explicit guidance, while reinforcement learning involves agent learning decisions to maximize rewards in an environment [12] [13]. Supervised learning's versatility finds applications in various domains, offering solutions in classification, regression, and ranking [14] [15].

1) Support Vector Machine (SVM): Stands out as a prominent supervised learning algorithm, initially crafted by Vapnik for binary classification and regression [37]. Renowned for its robust theoretical foundations, SVM has evolved to address multi-class classification using techniques like one-vs-rest and one-vs-one approaches. In the one-vs-rest method, distinct SVM models are trained for each class, treating it as positive and the others as negative, with the final result determined by the most probable classifier. Conversely, the one-vs-one strategy involves SVM models comparing each class against every other class, employing a voting scheme for the outcome. SVM excels in handling both linearly and non-linearly separable data. For linearly separable data, the hyperplane equation is defined as $g(x) = w^T x + b$, where w is the weight vector, x is the input data vector, and b is the bias term. The Euclidean norm determines the vector's magnitude, which is crucial for understanding its length in n-dimensional space. The optimization goal of SVM is to find the optimal hyperplane, maximizing the margin between support vectors accomplished through convex quadratic programming. SVM employs the kernel trick for non-linearly separable data, transforming input data into a higher-dimensional feature space for linear separation. The choice of the kernel function, whether linear, polynomial, radial basis function (RBF) or HyperTangent, profoundly influences SVM's ability to capture intricate patterns and relationships in the data [38]. SVM's suitability for text classification stems from its generalization capabilities, adherence to the Structural Risk Minimization principle, capacity to incorporate prior knowledge, and superior performance to alternatives like k-nearest-neighbors (kNN).

2) Logistic regression (LR) is a widely employed supervised learning algorithm that is a statistical method for modeling the probability of a binary outcome based on predictor variables. LR recognizes vectors with variables in text classification, assesses coefficients for each input, and predicts text class as a word vector. This model measures the statistical significance of independent variables concerning probability, offering a potent means of modeling binomial outcomes. LR excels in text categorization, providing advantages like computing probability values instead of scores. The logistic function, or sigmoid function, characterizes the LR model's relationship between variables and the probability of the outcome. Ensuring predicted probabilities fall within the range of 0 and 1, the LR equation is P = 1/(1+e^(- (w+bX))). In training, LR estimates parameters (weights) w and b through maximum likelihood estimation, aiming to maximize the likelihood of observed data. LR is computationally efficient, easy to implement, and yields interpretable results with estimated coefficients. It accommodates numerical and categorical input features and extends to multi-class classification using strategies like one-vs-rest, where a separate LR model is trained for each class, determining the final prediction based on the highest probability among all models.

3) Decision Tree (DT): A machine-learning algorithm for classification and regression tasks. It operates by recursively partitioning the data based on the values of input features, ultimately leading to a decision regarding the target variable. The algorithm constructs a tree-like structure representing a sequence of decisions and their potential consequences. Each internal node of the tree corresponds to a test on a specific attribute, while each branch represents the outcome of the test. The tree's leaf nodes correspond to class labels or numerical values [16]. Decision trees are popular in machine learning due to their interpretability, simplicity, and ability to handle various data types, including categorical and numerical variables [17]. They find applications in multiple domains, including image processing, clinical practice, and financial analysis. Their inherent structure allows for an intuitive understanding of the decision-making process, making them valuable tools for extracting insights from data.

D. Covid-19 and Vaccinations

The COVID-19 pandemic, caused by the novel coronavirus SARS-CoV-2, has profoundly impacted societies worldwide [28]. It was first identified when the initial case was reported in Wuhan, China, on December 19, 2019 [29]. The World Health Organization (WHO) officially declared the global COVID-19 pandemic on March 11, 2020 [30]. This declaration marked a turning point in the international response to the virus, leading to widespread public health measures to curb its spread. Saudi Arabia recorded its first confirmed case of COVID-19 on March 2, 2020 [31] [32]. The country swiftly implemented various measures to combat the virus's transmission, including lockdowns and travel restrictions.

The introduction of COVID-19 vaccines marked a pivotal moment in the fight against the pandemic. Saudi Arabia approved the Pfizer-BioNTech vaccine on December 10, 2020 [33]. Registration for the vaccine in Saudi Arabia began on December 15, 2020 [34]. As vaccination efforts progressed, restrictions evolved, with announcements such as allowing only vaccinated individuals to enter certain buildings starting from August 1, 2021. The vaccine rollout continued to advance, with the commencement of the second vaccine dose administration on June 23, 2021 [35]. As the situation improved, Saudi Arabia took steps to return to normalcy, including lifting many precautionary measures on May 3, 2022 [36].

III. Material and Method

This section provides a comprehensive overview of our methodology for producing a machine learning (ML) model and constructing a labeled corpus for emotion analysis in healthcare, specifically focusing on the Saudi dialect. The process comprises distinct stages, outlined in Fig. 1. The initial step involves collecting relevant tweets on vaccinations through specific keywords using the Twitter API. These tweets are then manually annotated by three Saudi natives, forming a labeled corpus for training and evaluation. Data preprocessing is undertaken to filter irrelevant content, rectify text errors, ensure consistency, and tokenize the text for analysis. Feature extraction follows, where we employ Bag-of-Words, N-Gram, and TF-IDF methods to effectively capture emotions expressed in the tweets, serving as inputs for the ML model. The classification stage involves developing and training the ML model, utilizing Support Vector Machines, Logistic Regression, and Decision trees to classify tweets into seven emotions. The performance evaluation includes using various metrics and experiments to validate the methodology. Unseen data is employed for testing, with metrics like accuracy, precision, recall, and F1-score measured to assess the effectiveness and limitations of our approach.

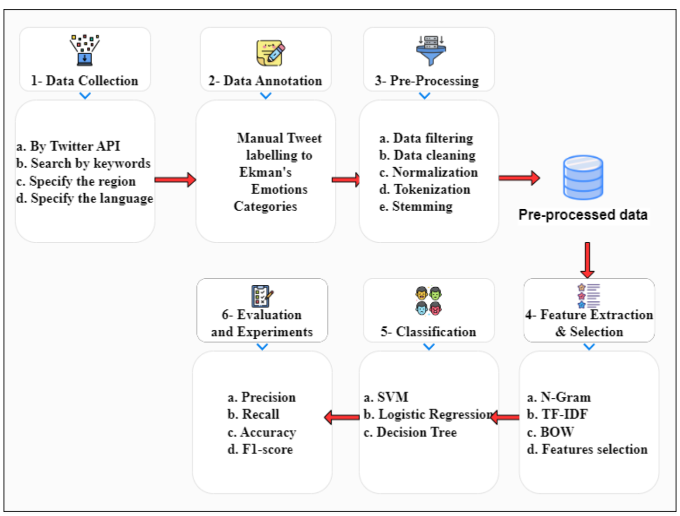


Fig. 1. Block diagram of the research methodology.

A. Dataset Collection

The data collection phase is pivotal to our implementation, aiming to amass diverse tweets related to vaccinations in the Saudi dialect. We utilized the Twitter API for Academic Research with Python programming language to access real-time Twitter data, employing libraries such as Tweepy [39], Pandas, and Requests. The collection spanned from December 15, 2020, to March 10, 2022, crucial periods in Saudi Arabia's COVID-19 vaccination timeline. Specific Arabic keywords targeting vaccination-related discussions were specified in each request, ensuring relevance. Additional parameters refined the dataset, focusing on the Saudi region and Arabic language and excluding retweets. The process yielded 34,074 raw tweets, each characterized by properties like tweet text, author ID, creation time, geolocation, and engagement metrics. The dataset, saved in CSV format, forms a robust foundation for subsequent annotation, preprocessing, feature extraction, and classification stages. Table I shows one raw instance of collected tweets.

TABLE I. One Raw Instance of the Collected Tweets

Tweet	@_doje_ @FBasmer انتهت حفلة كورونا!! وقطاع السوبرماركت
Author_ID	534798407
Created_at	2020-12-19 17:37:51+00:00
Geo	000799c66e428a87
ID	1.34035E+18
Lang	Ar
Like_count	0
Quote_count	0
Reply_count	0
Retweet_count	0
Source	Twitter for Android

B. Data Annotation

In the data annotation phase, emotions were manually assigned to collected tweets following initial preprocessing steps. These steps included eliminating duplicates, Twitter handles, URLs, English text, and emojis/non-emoji symbols. The annotation categorized tweets into emotion classes aligned with Paul Ekman's emotions [40], augmented by a neutral/spam class and an optimistic class to account for Saudi dialect nuances. Eight emotion categories were established, detailed in Table II.

TABLE II. Shows the Emotions Used in the Annotation Stage

#	English Emotion Class	Arabic Emotion Class	Explanation
1	happiness	سعادة	When the tweet expresses feelings of joy, happiness, or delight
2	fear	خوف	When the tweet expresses feelings of fear
3	disgust	اشمئزاز	when the tweet expresses disgust, disgust, or disgust with the tweeter
4	anger	غضب	When the tweet expresses angry feelings
5	surprise	تفاجؤ	When the tweet expresses feelings of surprise, astonishment, or wonder
6	optimism	تفاؤل	When the tweet expresses feelings of optimism and a positive view of the future
7	sadness	حزن	When the tweet expresses feelings of sadness, brokenness, grief, or depression
8	neutral	محايد/إعلان/لا يمكن تحديده	When the content of the tweet does not include any emotions or feelings that cannot be identified from other options or includes only advertising hashtags without any emotions

Google Sheets were employed in the annotation stage, where a copy of tweets was shared with individual annotators familiar with the Saudi dialect. Annotators were guided by detailed instructions ensuring privacy, single emotion selection for ambiguous cases, accuracy, and time management. The annotation process lasted two months, maintaining a manageable daily average for annotators. Results were consolidated into a dataset, revealing 23,689 fully matched tweets and 9,684 with discrepancies. Table III illustrates counts for each emotion class by annotators, and Fig. 2 provides examples of annotated tweets. This annotation phase produced a labeled dataset, a crucial foundation for subsequent stages like data preprocessing, feature extraction, and classification.

C. Preprocessing

The preprocessing stage, integral to our methodology, comprised two pivotal sub-stages: pre-annotation and post-annotation. In the pre-annotation phase, executed within Google Sheets using sophisticated REGEX formulas, we undertook various measures to refine the dataset comprehensively. Initially, we focused on eliminating redundancy, ensuring uniqueness in our dataset by removing 701 duplicate tweets from the initial count of 34,074, resulting in 33,373 distinct tweets. Furthermore, we removed Twitter handles (@), URLs, English text, emojis, and non-emoji symbols to enhance the dataset's purity. The cleaning tasks aimed at centering our analysis on the core content of tweets, devoid of any external influences. In the post-annotation, we employed the KNIME [41] platform to undertake further profound preprocessing exclusively on the subset of class-matched tweets from the three raters, totaling 22,689 tweets. Fig. 3 demonstrates the comprehensive nature of our data refinement process.

The post-annotation preprocessing involved a multifaceted approach:

- Punctuation Removal: Punctuation marks were expunged from tweet text to eliminate unnecessary noise, allowing focused analysis of words and their emotional significance.
- Elimination of Numbers: Numeric characters were systematically removed from tweets as they do not contribute directly to emotional content. This step simplified the text and heightened subsequent analysis accuracy.

TABLE III. COUNTS TWEETS ANNOTATED IN EACH EMOTION CLASS BY THREE ANNOTATORS

Rater		8	7	6	5	4	3	2	1
1#	#	21992	1723	1169	756	1358	821	1277	4277
	%	65.90%	5.16%	3.50%	2.27%	4.06%	2.47%	3.82%	12.82%
2#	#	26491	794	1177	19	1723	179	576	2414
	%	79.38%	2.38%	3.53%	0.06%	5.16%	0.53%	1.73%	7.23%
3#	#	22415	1313	1069	419	1499	840	1214	4604
	%	67.16%	3.93%	3.20%	1.25%	4.49%	2.51%	3.63%	13.79%

Tweet	class by rater #1	class by rater #2	class by rater #3
مدينة الضباب تغلق ابوابها ابتداء من الغد ديسمبر بسبب السلاله الجديده المتحوره من يروس كورونا اغلاق تام للبوتيكات والمطاعم والصالونات وجميع المحلات الغير اساسيه فقط الصيدليات والسوبر ماركات بتكون فاتحه	5 (surprise)	8 (neutral/spam)	1 (happiness)
الحمدلله اخذت اول جرعة من اللقاح اللهم انفع بها خذ الخطوة	1 (happiness)	1 (happiness)	1 (happiness)
تعبت نفسيا والسبب ان كورونا للحين مو راضي يخلص	7 (sadness)	8 (neutral/spam)	2 (fearness)
اذا شركة واحدة اللي بتصنع اللقاح بيكون احتكار والشركات الطبية كل سنة تزيد توسعها علشان الدواء مثل له اكثر من شركة مصنعة هو نفس الحال مع كورونا	8 (neutral/spam)	8 (neutral/spam)	8 (neutral/spam)

Fig. 2. Examples of annotated tweets by different raters.

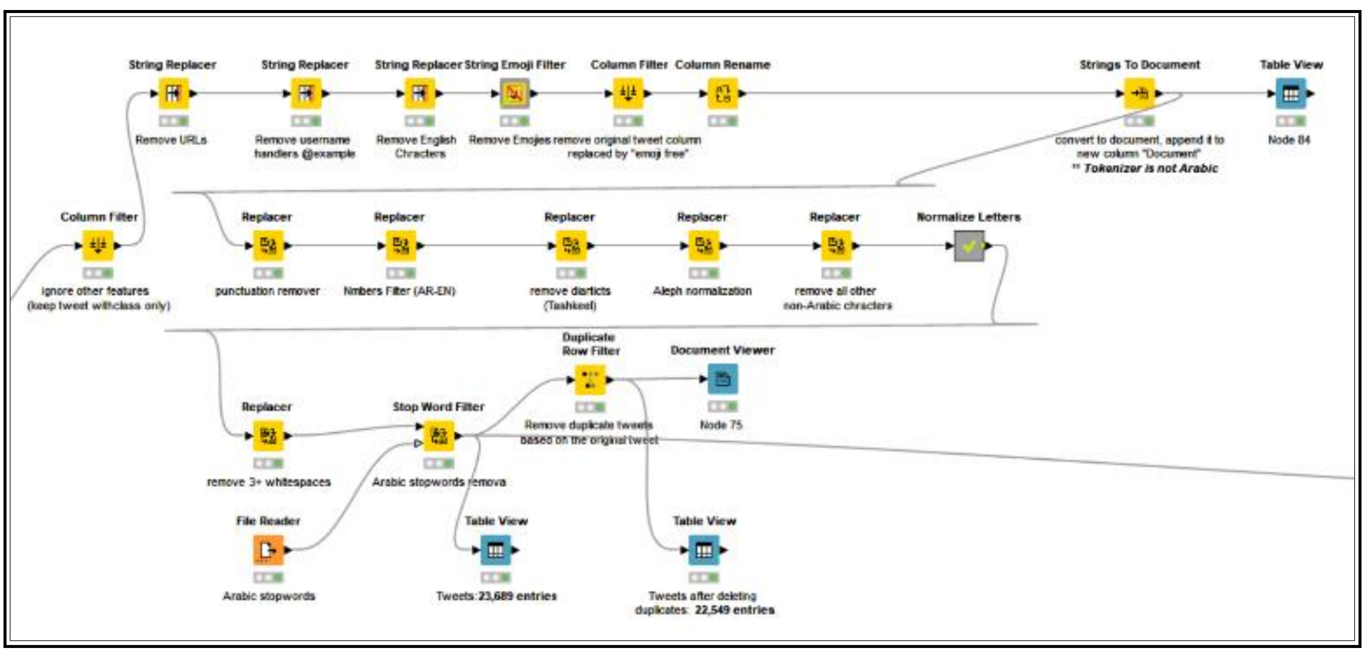
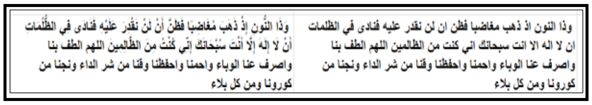


Fig. 3. The flow of preprocessing nodes in the KNIME platform.

- Removal of Double Spaces and New Lines: Consecutive spaces and new lines were eradicated, ensuring uniformity in text format.

- Arabic Diacritics Removal and Character Normalization: Diacritics, such as vowel marks, were deleted for consistency. Characters were normalized to ensure uniformity and standardization across the dataset. This eliminated text variations that could impact emotion analysis accuracy, as shown in Fig. 4.



وَذَا النُّون إِذْ ذَهَبَ مُغَاضِبًا فَظَنَّ أَنْ لَنْ نَقْدِرَ عَلَيْهِ فَنَادَى فِي الظُّلُمَاتِ أَنْ لَا إِلَهَ إِلَّا أَنْتَ سُبْحَانَكَ إِنِّي كُنْتُ مِنَ الظَّالِمِينَ اللهم الطف بنا واصرف عنا الوباء واحمنا واحفظنا وقنا من شر الداء ونجنا من كورونا ومن كل بلاء	وذا النون اذ ذهب مغاضبا فظن ان لن نقدر عليه فنادى في الظلمات ان لا اله الا انت سبحانك اني كنت من الظالمين اللهم الطف بنا واصرف عنا الوباء واحمنا واحفظنا وقنا من شر الداء ونجنا من كورونا ومن كل بلاء

Fig. 4. An example of a tweet before and after diacritics removal.

- Normalization Techniques: Normalization techniques were applied to standardize data and convert characters to their original shape. For instance, sequences of certain Arabic letters were regulated for uniformity.

- Stemming: Employing three distinct stemming techniques (Snowball Stemmer, Porter Stemmer, and Kuhlen Stemmer), words were reduced to their root form. This facilitated a more accurate analysis and interpretation of emotions.

- Stopwords that lack significant meaning were removed using a Stop Word Filter node in KNIME. A stop-word dictionary for the Arabic language was employed for this purpose [42].

- Tokenization: Text was tokenized into individual words using the Arabic tokenizer provided by the NLTK library.

- Column Filtering: Irrelevant columns were filtered out, retaining only tweet and label class columns, streamlining subsequent emotion analysis.

This meticulous preprocessing led to a pristine dataset comprising 22,549 clean tweets. This refined dataset is now poised for the subsequent stages of feature extraction, classification, and further analysis, as shown in Fig. 5.

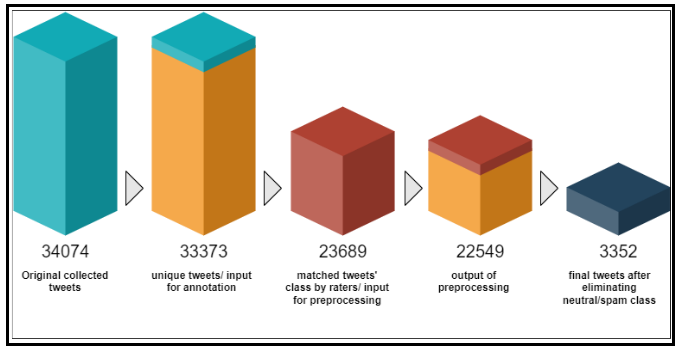


Fig. 5. Counts the tweets at each stage of the implementation.

D. Feature Extraction and Selection

In the feature extraction and selection stage, our initial step involved the removal of the 8th emotion class label, encompassing neutral, spam, and advertisement text (totaling 19,197 tweets). This exclusion aimed to streamline subsequent analysis, focusing solely on the emotions inherent in the remaining dataset of 3,352 tweets.

1) Bag-of-Words (BoW): The Bag-of-Words approach, a fundamental yet robust technique, was employed to convert preprocessed text data into numerical vectors. Each word in the text is treated as a distinct feature, and its frequency is quantified in this method.

The resulting vector encapsulates the occurrence of each word in the text, irrespective of word order or grammar. Two sub-flows of BoW were applied to evaluate model accuracy, each depicted in. This method yielded a high-dimensional representation of the text data, forming the input for subsequent machine-learning models. Fig. 6 and Fig. 7 show these flows.

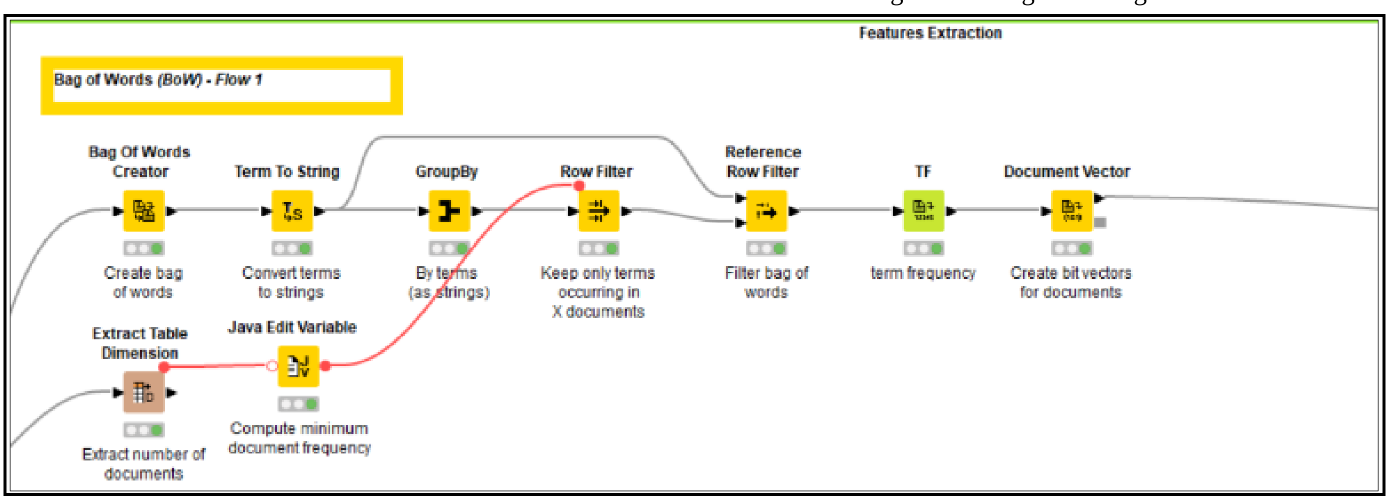


Fig. 6. First sequence of BoW flow.

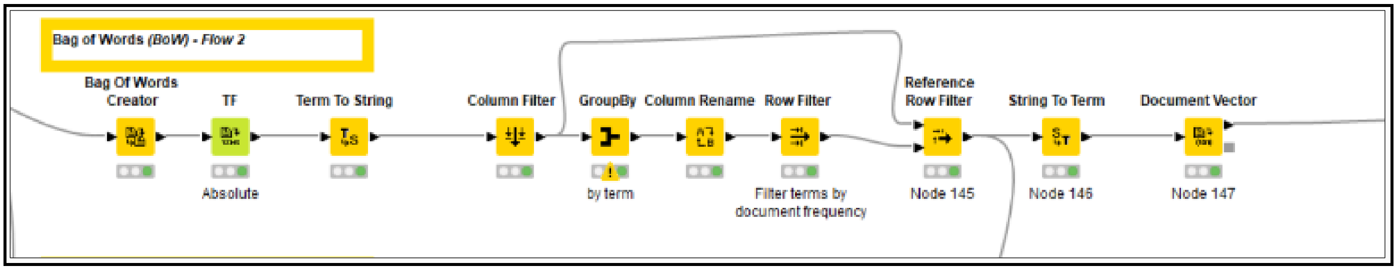


Fig. 7. Second sequence of BoW flow.

2) N-Gram: Extending the Bag-of-Words approach, the N-Gram method considers sequences of N consecutive words as features. We specifically employed bi-grams (N=2), representing pairs of successive words in each tweet. This extension allows for capturing context and word relationships in the text, enhancing our ability to discern nuanced insights into expressed emotions.

3) Term Frequency-Inverse Document Frequency (TF-IDF): TF-IDF, a prevalent technique in natural language processing, gauges the importance of each word in a document relative to a corpus. It factors in the frequency of a word in a specific document and its rarity across the entire corpus. Assigning higher weights to words that are frequent in a document but rare across the corpus makes them more discriminative. TF-IDF was employed to enrich the representation of words based on their significance in each tweet and across the entire dataset.

We transformed the preprocessed text data into numerical representations using these three feature extraction methods. These representations effectively captured essential information about the emotions expressed in the tweets. The resulting feature matrices served as inputs for the machine learning model during training, facilitating the model's ability to discern patterns and relationships between features and labeled emotions.

E. Resolve Data Imbalance (Oversampling)

The focus is on addressing class imbalance, a crucial consideration in developing an effective emotion analysis model. The section outlines the strategies employed, particularly data partitioning and oversampling techniques.

1) Class distribution: An evaluation of the initial class distribution within the dataset is conducted before diving into oversampling. It's revealed that the original dataset of 3,352 tweets exhibits an imbalance across emotion categories, a factor that can impact the model's ability to discern less frequent emotions effectively as shown in Table IV and Fig. 8.

TABLE IV. CLASS DISTRIBUTIONS OF THE 3352 TWEETS BEFORE DATA SPLITTING

Class	Tweets Count
Happiness (1)	1,926
Fear (2)	230
Disgust (3)	71
Anger (4)	482
Surprise (5)	13
Optimism (6)	243
Sadness (7)	387
Total	**3352**

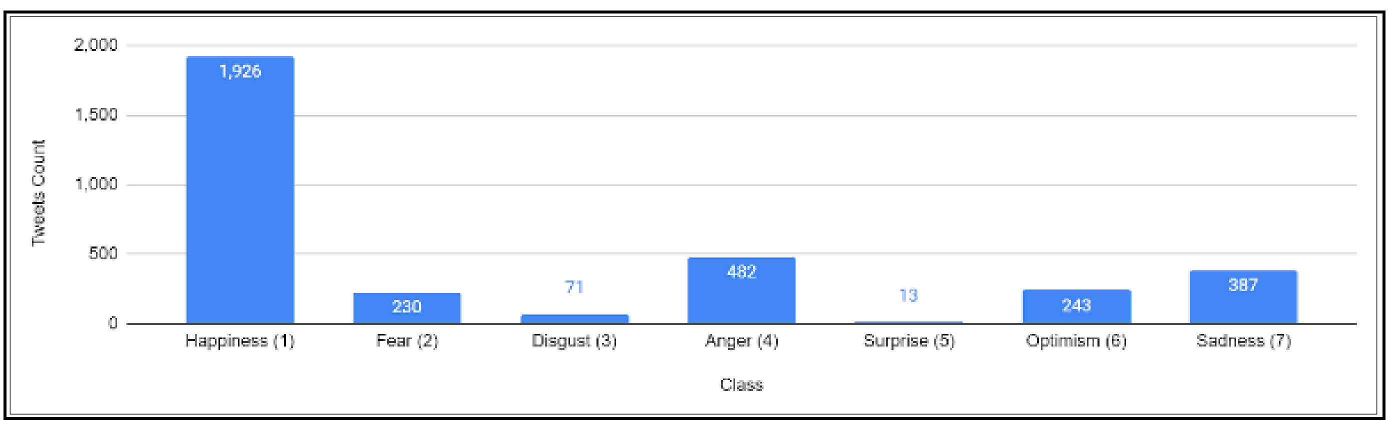


Fig. 8. Class distributions of the 3352 tweets before data splitting.

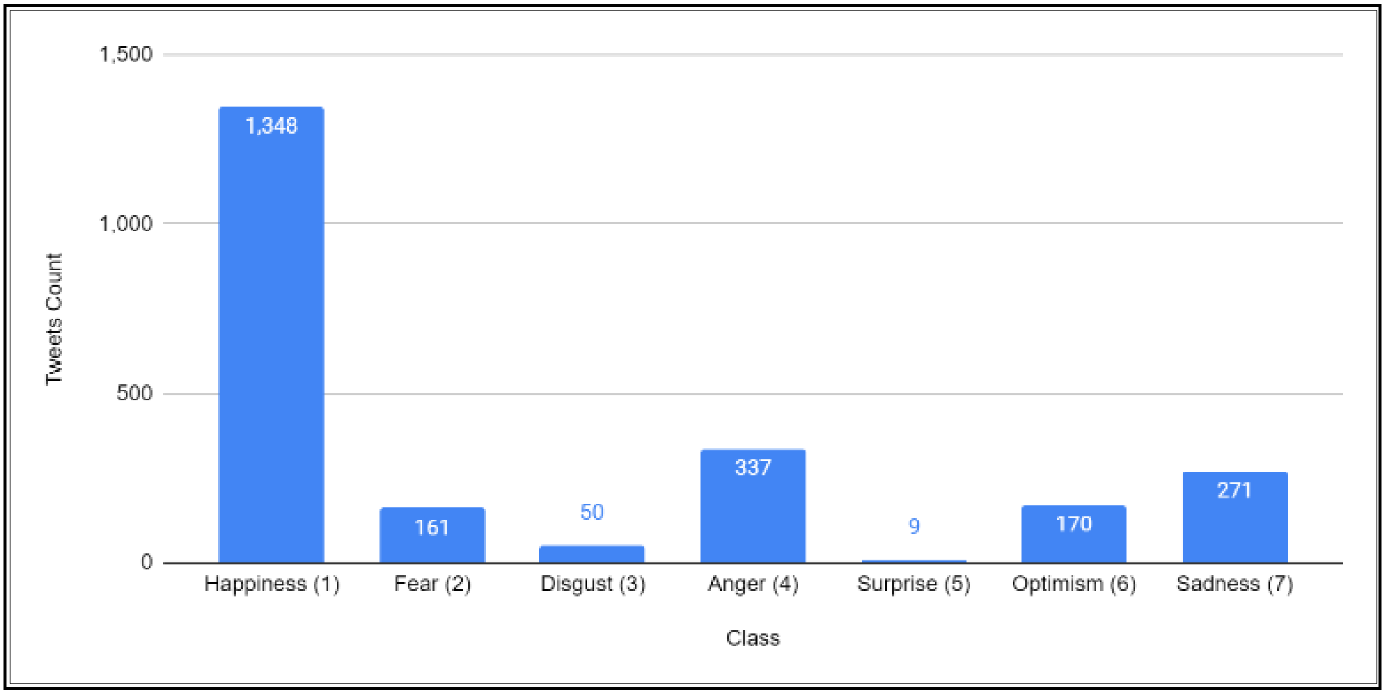


Fig. 9. Class distribution of the training dataset contains 2346 tweets after data splitting and before oversampling.

2) Dataset splitting: A 70:30 train-test split ratio is employed to evaluate machine learning model performance. This ensures that 70% of the data is allocated for training, and the remaining 30% is reserved for evaluation. The class distribution post-splitting is presented for both the training and testing datasets as shown in Table V and Fig. 9.

TABLE V. THE CLASS DISTRIBUTION OF THE TRAINING DATASET CONTAINS 2346 TWEETS AFTER DATA SPLITTING AND BEFORE OVERSAMPLING

Class	Tweets Count
Happiness (1)	1,348
Fear (2)	161
Disgust (3)	50
Anger (4)	337
Surprise (5)	9
Optimism (6)	170
Sadness (7)	271
Total	**2346**

3) Oversampling: Oversampling techniques are strategically applied to rectify the class imbalance within the training data. The oversampling is exclusively directed at the training data. After this process, the class distribution in the training dataset is significantly altered, as illustrated in Table VI and Fig. 10.

TABLE VI. CLASS DISTRIBUTION OF THE TRAINING DATASET CONTAINS 9399 TWEETS AFTER DATA SPLITTING AND AFTER OVERSAMPLING

Class	Tweets Count
Happiness (1)	1,348
Fear (2)	1288
Disgust (3)	1350
Anger (4)	1348
Surprise (5)	1350
Optimism (6)	1360
Sadness (7)	1355
Total	**9399**

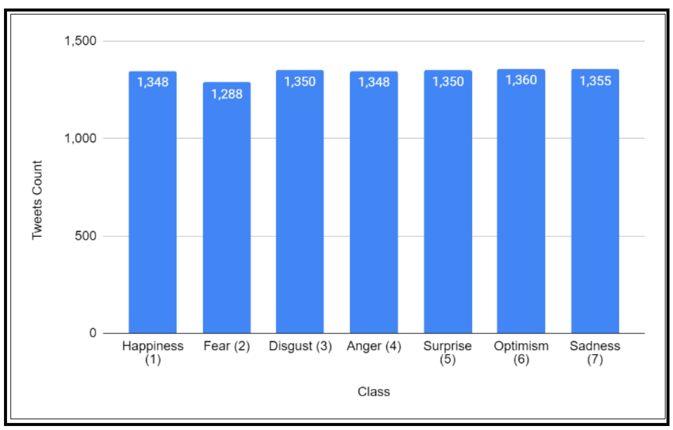


Fig. 10. Class distribution of the training dataset contains 9399 tweets after data splitting and after oversampling.

This oversampling endeavor aims to equalize the representation of emotions across classes in the training dataset, providing a more balanced learning experience for machine learning models.

IV. Implementation

In the Classification stage, we embarked on the crucial task of predicting the emotions expressed in the 9399 tweets using the extracted features. Indeed, the combination of three stemming techniques, four feature extraction methods, and three machine learning algorithms resulted in 36 different models being trained in this stage. Each model represents a unique configuration of the preprocessing and classification pipeline, contributing to the comprehensive evaluation and comparison of various approaches. We obtained an enriched dataset with numerical representations of the extracted features. However, the emotion labels must be converted to numerical classes to train the machine learning models. This transformation involved mapping each emotion category to a unique numerical class, making the dataset suitable for classification. Having prepared the data and set up the train-test split, we applied three widely used machine learning algorithms: Support Vector Machine (SVM), Logistic Regression, and Decision Tree. Each algorithm underwent training on the training data to learn the underlying patterns and relationships between the extracted features and emotions. After training, the models were tested using the test data to evaluate their predictive capabilities.

In the learner node of each machine learning algorithm, specific configurations and options were carefully selected to optimize the performance of the models. For the Support Vector Machine (SVM), we utilized the polynomial kernel with power approximately equal to 1, bias around 1, and gamma set to around 1. The overlapping penalty was established within the range of 0.1 to 1, enabling us to control the influence of overlapping data points on the model's decision boundaries. In the case of logistic regression, we opted for the stochastic average gradient (SAG) solver, known for its efficiency [25] and ability to handle large datasets. The maximal number of epochs was set between 10 and 20, ensuring the algorithm converged to the optimal solution while avoiding overfitting. For the Decision Tree algorithm, we set the maximum number of patterns the tree will store to support highlighting to a default value of 10,000. This configuration allowed us to manage the complexity of the tree while still capturing the relevant patterns and relationships in the data flows illustrated.

V. Evaluation and Results

In this section, we will evaluate the model results using different metrics. We utilized the KNIME platform for the evaluation process, employing the Model Predictor and Scorer nodes. The Model Predictor node takes the test data partition from the partitioning node and the trained model from the learner node as inputs. It then uses the trained model to predict the emotion labels for the given test data. The Scorer node compares the predicted and actual labels, generating a confusion matrix displaying correct and incorrect predictions for each emotion class.

A. Evaluation Methods

1) Confusion matrix: a fundamental evaluation tool that provides a tabular representation of the performance of a machine-learning classification model [26]. It gives a comparison between actual and predicted values as it displays the number of true positive (TP), true negative (TN), false positive (FP), and false negative (FN) predictions for each emotion class, which will be used in the measure. Furthermore, it allows us to assess the model's accuracy and ability to classify emotions correctly. The confusion matrix is a square matrix of size N x N, where N represents the number of emotion classes. We have a 7 x 7 confusion matrix in our specific case, as illustrated in Table VII.

TABLE VII. Confusion Matrix for Seven Classes

Actual Class	**Happiness**	**Fear**	**Disgust**	**Anger**	**Sadness**	**Suprise**	**Optimism**
Happiness	TP1	FP1	FP2	FP3	FP4	FP5	FP6
Fear	FP7	TP2	FP8	FP9	FP10	FP11	FP12
Disgust	FP13	FP14	TP3	FP15	FP16	FP17	FP18
Anger	FP19	FP20	FP21	TP4	FP22	FP23	FP24
Sadness	FP25	FP26	FP27	FP28	TP5	FP29	FP30
Suprise	FP31	FP32	FP33	FP34	FP35	TP6	FP36
Optimism	FP37	FP38	FP39	FP40	FP41	FP42	TP7

2) Class distribution consideration: As our dataset exhibits an imbalanced class distribution, we applied stratified cross-validation with some emotion classes having significantly fewer instances than others. Stratified cross-validation ensures that each fold retains the same proportion of instances for each emotion class as the original dataset. This approach is crucial for preventing biased evaluations and ensuring that each emotion class is represented appropriately during model training and testing.

3) Comparative analysis: To conduct a comparative analysis, we evaluated a total of 36 different models, considering the combination of three stemming techniques (Kohlen Stemmer, Porter Stemmer, and Snowball Stemmer), four feature extraction methods (Bag-of-Words, N-Gram, and TF-IDF), and three machine learning algorithms (Support Vector Machine, Logistic Regression, and Decision Tree). By comparing the performance of these models, we can identify the most practical combination of techniques and algorithms for sentiment analysis in the context of our research.

B. Evaluation of Performance Metrics

A comprehensive set of performance metrics is employed to evaluate the effectiveness of emotion classification models for Saudi dialect tweets related to vaccinations. The chosen metrics, including Accuracy, Precision, Recall, and F1-score, offer valuable insights into the models' ability to classify emotions accurately.

1) Accuracy: a widely used metric in classification models, measures the proportion of correctly classified instances over the total number of instances in the dataset. It provides an overall assessment of the model's performance in correctly identifying emotions in tweets across all emotion classes. The formula for accuracy involves True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN).

2) Precision: assesses the proportion of true positive predictions for a specific emotion class over the total number of instances predicted as that class. It signifies the model's reliability in avoiding false positive predictions for a given class. Precision is crucial in understanding the model's accuracy when classifying tweets as a particular emotion.

3) Recall: sensitivity, or true positive rate, measures the proportion of true positive predictions for a particular emotion class over the total number of instances belonging to that class. It indicates the model's effectiveness in correctly identifying all instances of a specific emotion class. Recall is significant in understanding how well the model captures tweet emotions.

4) F1-score: a harmonic mean of precision and recall that combines both metrics into a single value. This score is instrumental in scenarios with an imbalance in class distribution. It offers a balanced evaluation of the model's performance, considering trade-offs between precision and recall in emotion classification.

C. Results and Evaluation

This section presents the outcomes of the 36 experiments conducted during the implementation stage. Table VIII, IX and Table X provide a comprehensive breakdown of the results for all 36 models, giving the performance metrics for each emotion class. These tables offer a basis for comparing each model's accuracy, precision, recall, and F1-score of each model across different feature extraction methods and stemming techniques. Among the configurations, the Logistic Regression model achieved the highest accuracy, reaching 74.95% when combined with N-Gram feature extraction and Snowball stemming; it also performs the same 74.95% accuracy when combined with the BoW with Snowball stemmer, followed by 74.75% Logistic Regression when combined with TF-IDF and Snowball Stemmer. The SVM model showcased a close performance with 74.35% accuracy when combined with N-Gram and Snowball.

Moving on to the recall metric, the SVM model achieved a remarkable 91.34% in both experiments, compromising N-Gram and TF-IDF when combined with the Snowball stemmer, displaying a higher ability to identify true positive instances correctly and making them top performers in emotion classification. The Logistic Regression result shows a close percentage of 91.00% with TF-IDF and Snowball stemmer.

Regarding precision, the SVM model demonstrated an impressive 98.98% precision rate when trained after using the N-Gram technique and Porter stemmer, indicating its capability to limit the number of false positives and ensure the accuracy of positive predictions. The Logistic Regression result shows a close percentage of 97.23% with BoW and Kuhlen Stemmer, followed by another Logistic Regression experiment with Bow and Porter Stemmer achieving 96.60%.

The F-measure results highlighted the Logistic Regression model as the most balanced performer between precision and recall, achieving 92.93% and 92.51%, particularly when combined with TF-IDF and BoW feature extraction techniques with the Snowball stemmer. The SVM model followed closely in third place, achieving 92.38% when N-Gram and Snowball were used.

TABLE VIII. SVM MODEL RESULTS ARE BASED ON FOUR FEATURE EXTRACTION TECHNIQUES AND THREE STEMMING METHODS

	Stemming	Accuracy %	Recall %	Precision %	F-Measure %
BoW-1	**Kuhlen**	64.31	83.04	95.61	88.88
	Snowball	67.39	85.46	96.10	90.47
	Porter	64.61	84.94	95.71	90.00
BoW-2	**Kuhlen**	70.67	87.54	93.35	90.35
	Snowball	74.15	89.96	94.37	92.11
	Porter	70.57	87.02	93.32	90.06
N-Gram	**Kuhlen**	70.67	87.19	92.98	90.00
	Snowball	74.35	91.34	93.45	92.38
	Porter	70.17	87.19	98.98	90.00
TF-IDF	**Kuhlen**	70.17	87.19	92.98	90.00
	Snowball	74.35	91.34	93.45	92.38
	Porter	70.17	87.19	92.98	90.00

TABLE IX. DECISION TREE MODEL RESULTS ARE BASED ON FOUR FEATURE EXTRACTION TECHNIQUES AND THREE STEMMING METHODS

	Stemming	Accuracy %	Recall %	Precision %	F-Measure %
BoW-1	**Kuhlen**	67.29	83.73	95.46	89.21
	Snowball	66.10	84.25	92.40	88.14
	Porter	67.29	83.73	95.46	89.21
BoW-2	**Kuhlen**	66.00	84.42	94.39	89.13
	Snowball	69.38	88.23	90.58	89.39
	Porter	66.79	86.85	91.27	89.00
N-Gram	**Kuhlen**	66.79	86.85	91.27	89.00
	Snowball	69.38	88.23	90.58	89.39
	Porter	66.79	86.85	91.27	89.00
TF-IDF	**Kuhlen**	65.90	84.25	95.49	89.52
	Snowball	67.99	85.81	92.36	88.96
	Porter	66.00	84.42	94.39	89.13

D. Discussion

The results from our comprehensive experiment illuminate the effectiveness of our proposed Emotion Analysis machine learning model in classifying emotions and feelings expressed in Tweets written in the Saudi dialect. Our study's primary aim was to develop a precise and dependable model that enhances existing study results and aligns with the Saudi dialect's unique linguistic and cultural intricacies. As demonstrated in Fig. 11 to Fig. 14. The most accurate, precise recall and F1-score models are highlighted. The Logistic Regression model, as depicted, surpasses others in terms of Accuracy and F-Measure metrics, signifying a significant achievement. The SVM model displays the highest Recall (sensitivity) and Precision performance. The third model (Decision Tree) is considered out of the competition of the top three. This success contributes to the validation of our objectives.

TABLE X. SHOWS THE LOGISTIC REGRESSION MODEL RESULTS BASED ON FOUR FEATURE EXTRACTION TECHNIQUES AND THREE STEMMING METHODS

	Stemming	**Accuracy %**	**Recall %**	**Precision %**	**F-Measure %**
BoW-1	**Kuhlen**	63.81	79.23	97.23	87.32
	Snowball	69.28	87.19	94.73	69.28
	Porter	66.00	83.73	96.60	89.71
BoW-2	**Kuhlen**	72.86	89.79	93.68	91.69
	Snowball	74.95	90.83	94.25	92.51
	Porter	73.36	89.61	93.84	91.68
N-Gram	**Kuhlen**	73.06	89.96	94.71	92.28
	Snowball	74.95	90.31	94.05	74.95
	Porter	73.26	89.10	94.49	91.71
TF-IDF	**Kuhlen**	72.16	87.88	94.24	90.95
	Snowball	74.75	91.00	94.94	92.93
	Porter	72.76	88.58	93.60	91.02

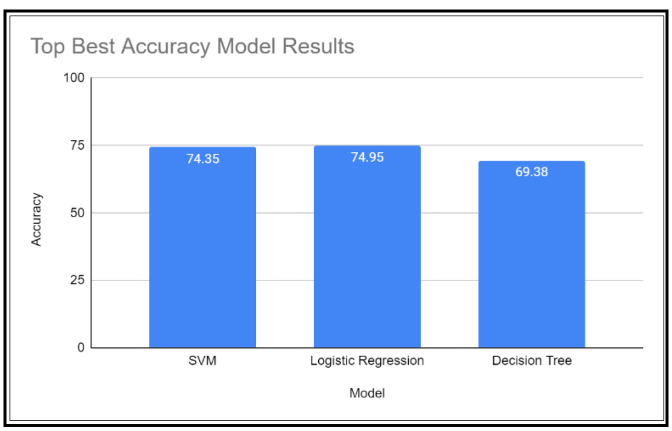


Fig. 11. The top accuracy results achieved by the three models.

When compared to existing works in the domain, our progress is remarkable. For instance, considering a study that achieved 73.39% accuracy through an SVM approach, which is 1.56% lower than our Logistic Regression result, and they solely employed two machine learning algorithms (SVM and MNB) [11], our accomplishments are striking. Not only did we surpass this accuracy benchmark, but we also encompassed a diverse array of machine-learning algorithms, leading to a more robust and comprehensive evaluation. This also aligns with our objectives.

The performance of our model was significantly influenced by the careful selection of stemming techniques and feature extraction methods. Results indicate that specific combinations, such as employing N-Gram feature extraction with Snowball stemming, yield the highest accuracy. This underscores the importance of selecting the correct machine learning algorithm and optimizing preprocessing and feature engineering stages to exploit the data's potential fully.

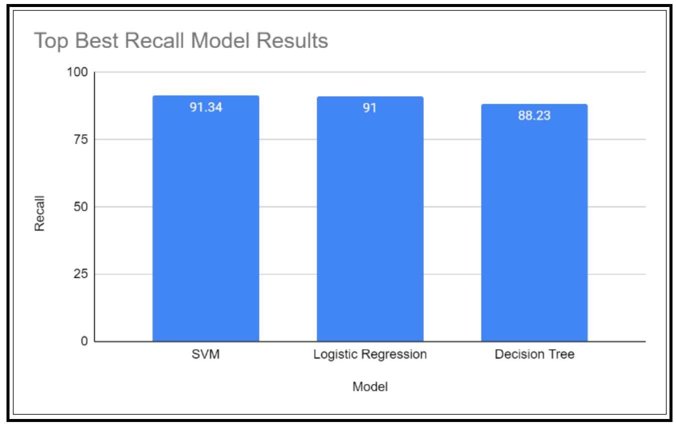


Fig. 12. The top recall results achieved by the three models.

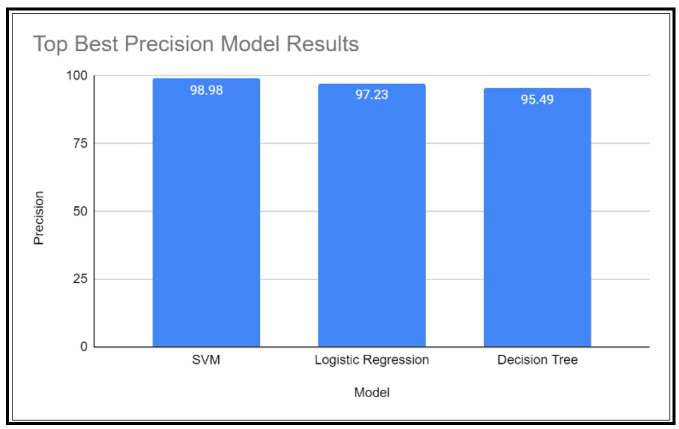


Fig. 13. The top precision results achieved by the three models.

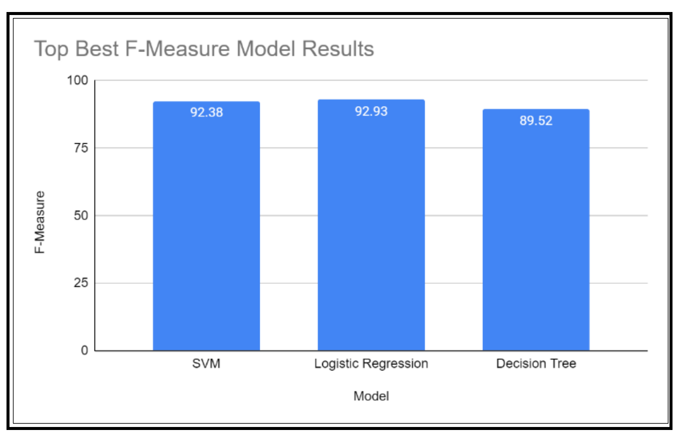


Fig. 14. Shows the top F-Measure Results achieved by the three models.

E. General Attitudes Towards COVID-19 Vaccinations in Saudi Arabia

The outcomes of our model shed light on prevalent concerns and sentiments regarding COVID-19 vaccination in Saudi Arabia. The precision of our emotion analysis allows us to extract insightful understandings from public sentiment, aiding decision-making in healthcare initiatives. The distribution of annotated tweets across emotion classes is detailed in Table XI. They are revealing a diverse emotional landscape. 'Happiness' dominates, followed by 'Anger,' 'Sadness,' and 'Optimism.' 'Disgust' is rare, and 'Surprise' is infrequent. Insights and Latent Dirichlet Allocation (LDA) [27] indicate a prevailing positive disposition toward the COVID-19 vaccination campaign in Saudi Arabia. As shown in Fig. 15, 57.46% of tweets expressed happiness followed by anger 14.38%, sadness 11.55%, and optimism 7.25%. Fearful emotions account for 6.86%, while disgust and surprise are 2.12% and 0.39%, respectively. These findings highlight contentment with vaccine availability, achieving our objective of providing valuable insights into general attitudes toward vaccinations.

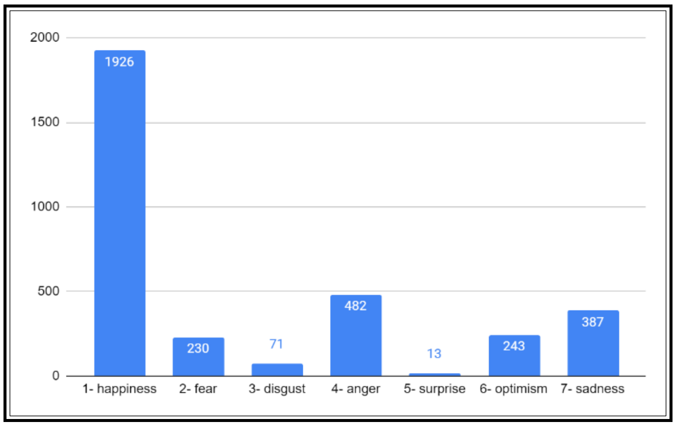


Fig. 15. The distribution of annotated tweets across various emotion classes.

TABLE XI. NUMBER OF INSTANCES IN EACH EMOTION CLASS IN THE FINAL STAGE OF IMPLEMENTATION

Class	Number of Instances	%
1- happiness	1926	57.5
2- fear	230	6.9
3- disgust	71	2.1
4- anger	482	14.4
5- surprise	13	0.4
6- optimism	243	7.2
7- sadness	387	11.5

F. Saudi Dialect Labeled-Tweets Corpus Availability

In alignment with our objective to produce a Saudi dialect labeled-tweets corpus in the healthcare and COVID-19 vaccination domain, we are pleased to announce the availability of the "saudiEAR" repository on GitHub [43]. This repository contains a comprehensive collection of tweets, including both the original dataset collected and the preprocessed version. By making this corpus publicly accessible, we aim to contribute to the research community and facilitate advancements in sentiment analysis, emotion classification, and related fields. This open dataset enables researchers and practitioners to explore the intricacies of sentiment expression in the Saudi dialect, particularly in healthcare and COVID-19 vaccination discussions.

VI. CONCLUSION AND FUTURE WORK

A. Conclusion

In conclusion, this research addresses a notable gap in emotional analysis, focusing on the Saudi context amid COVID-19 vaccination discussions. The objective was to build an effective machine-learning model for classifying Saudi tweets into distinct emotions, a need prompted by the scarcity of such studies in the Saudi dialect and the absence of a suitably labeled corpus. The methodology involved a meticulous collection of 34,074 Arabic tweets emphasizing COVID-19 vaccines in Saudi Arabia. Three expert raters annotated these tweets into eight emotion classes, followed by thorough preprocessing, resulting in a dataset of 3,352 tweets expanded through oversampling to 9,399. Thirty-six machine learning experiments were conducted, employing SVM, Logistic Regression, and Decision Trees, with three stemming techniques and four feature extraction methods. Key findings reveal the Logistic Regression model achieving a noteworthy accuracy of 74.95%. The SVM model excelled with a 91.34% recall and 98.98% precision. The F1-Score for Logistic Regression reached 92.93%, showcasing the approach's effectiveness. Comparative analysis with existing studies indicated accuracy, precision, recall, and F1-Score improvements.

Insights from the dataset highlighted a prevailing positive sentiment toward COVID-19 vaccination campaigns. Happiness dominated at 57.5%, followed by anger (14.4%) and sadness (11.5%). These sentiments mirror people's joy for potential pandemic resolution, reflecting trust in government decisions. As a contribution, the labeled dataset of 33,373 tweets is provided, facilitating further research in emotion analysis within the Saudi dialect and supporting advancements in machine learning and sentiment analysis.

B. Future Work

Our study has opened many possibilities for future research. First, we need to improve the performance of our model by incorporating deep learning methods. More advanced neural network architectures, such as recurrent or transformer-based models, could allow us to make more nuanced sentiment classifications, especially when capturing context-dependent linguistic nuances. In addition, we can gain deeper insights into the linguistic and contextual cues that underlie the expression of emotions by exploring the interpretability of our model's decisions. Techniques such as attention mechanisms or layer-wise relevance propagation can help us understand how the model makes its decisions.

Finally, we can broaden the scope of emotion analysis and its implications for healthcare communication, public sentiment monitoring, and decision-making by extending our model's applicability to other dialects and languages within the Arab region.

REFERENCES

[1] Statista. (2023). Leading social networks worldwide as of January 2023, ranked by number of active users (in millions) [Graph]. In Statista. Retrieved April 25, 2023, from https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/

[2] Datareportal. (2023). Digital 2023 Saudi Arabia. Retrieved April 25, 2023, from https://datareportal.com/reports/digital-2023-saudi-arabia

[3] Adel Assiri, Ahmed Emam, and Hmood Al-Dossari. Towards enhancement of a lexicon-based approach for Saudi dialect sentiment analysis. Journal of Information Science, 44(2):184–202, April 2018.

[4] Rahma M. Alahmary, Hmood Z. Al-Dossari, and Ahmed Z. Emam. Sentiment Analysis of Saudi Dialect Using Deep Learning Techniques. In 2019 International Conference on Electronics, Information, and Communication (ICEIC), pages 1–6, Auckland, New Zealand, January 2019. IEEE.

[5] Erick Kauffmann, Jesús Peral, David Gil, Antonio Ferrández, Ricardo Sellers, Higinio Mora, A framework for big data analytics in commercial social networks: A case study on sentiment analysis and fake review detection for marketing decision-making, Industrial Marketing Management, Volume 90, 2020, Pages 523-537,ISSN 0019-8501, https://doi.org/10.1016/j.indmarman.2019.08.003. (https://www.sciencedirect.com/science/article/pii/S0019850118307612)

[6] Moghaddam, S. (2015). Beyond Sentiment Analysis: Mining Defects and Improvements from Customer Feedback. In: Hanbury, A., Kazai, G., Rauber, A., Fuhr, N. (eds) Advances in Information Retrieval. ECIR 2015. Lecture Notes in Computer Science, vol 9022. Springer, Cham. https://doi.org/10.1007/978-3-319-16354-3_44

[7] Han, J., Kamber, M., & Pei, J. (2012). Data mining: Concepts and techniques (3rd ed.). Morgan Kaufmann Publishers.

[8] Contreras-Valdes, A., Amezquita-Sanchez, J. P., Granados-Lieberman, D., & Valtierra-Rodriguez, M. (2020). Predictive Data Mining Techniques for Fault Diagnosis of Electric Equipment: A Review. Applied Sciences, 10(3), 950. https://doi.org/10.3390/app10030950

[9] Haifa K. Aldayel and Aqil M. Azmi. Arabic tweets sentiment analysis – a hybrid scheme. Journal of Information Science, 42(6):782–797, December 2016.

[10] Abdulmohsen Al-Thubaity, Mohammed Alharbi, Saif Alqahtani, and Abdulrahman Aljandal. A Saudi Dialect Twitter Corpus for Sentiment and Emotion Analysis. In 2018 21st Saudi Computer Society National Computer Conference (NCC), pages 1–6, Riyadh, April 2018. IEEE.

[11] Abeer A. AlFutamani and Heyam H. Al-Baity. Emotional Analysis of Arabic Saudi Dialect Tweets Using a Supervised Learning Approach. Intelligent Automation & Soft Computing, 29(1):89–109, 2021.

[12] Mahesh, B. (2020). Machine learning algorithms-a review. International Journal of Science and Research (IJSR).[Internet], 9, 381-386.

[13] Abhijit Gosavi, (2008) Reinforcement Learning: A Tutorial Survey and Recent Advances. INFORMS Journal on Computing 21(2):178-192. https://doi.org/10.1287/ijoc.1080.0305

[14] Domingos, P. (2012). A few useful things to know about machine learning. Communications of the ACM, 55(10), 78-87.

[15] Liu, T. Y. (2009). Learning to rank: From pairwise approach to listwise approach. Proceedings of the 24th International Conference on Machine Learning, 129-136.

[16] Charbuty, B., & Abdulazeez, A. (2021). Classification based on decision tree algorithm for machine learning. Journal of Applied Science and Technology Trends, 2(01), 20-28.

[17] Pranckevičius, T., & Marcinkevičius, V. (2017). Comparison of naive bayes, random forest, decision tree, support vector machines, and logistic regression classifiers for text reviews classification. Baltic Journal of Modern Computing, 5(2), 221.

[18] Sumayh S. Aljameel, Dina A. Alabbad, Norah A. Alzahrani, Shouq M. Alqarni, Fatimah A. Alamoudi, Lana M. Babili, Somiah K. Aljaafary, and Fatima M. Alshamrani. A Sentiment Analysis Approach to Predict an Individual's Awareness of the Precautionary Procedures to Prevent COVID-19 Outbreaks in Saudi Arabia. International Journal of Environmental Research and Public Health, 18(1):218, December 2020.

[19] Bador Al sari, Rawan Alkhaldi, Dalia Alsaffar, Tahani Alkhaldi, Hanan Almaymuni, Norah Alnaim, Najwa Alghamdi, and Sunday O. Olatunji. Sentiment analysis for cruises in Saudi Arabia on social media platforms using machine learning algorithms. Journal of Big Data, 9(1):21, December 2022.

[20] Lina A. Alhuri, Hutaf R. Aljohani, Rahaf M. Almutairi, and Fazilah Haron. Sentiment Analysis of COVID-19 on Saudi Trending Hashtags Using Recurrent Neural Network. In 2020 13th International Conference on Developments in eSystems Engineering (DeSE), pages 299–304, Liverpool, United Kingdom, December 2020. IEEE.

[21] Reem AlYami and Rabeah AlZaidy. Arabic Dialect Identification in Social Media. In 2020 3rd International Conference on Computer Applications & Information Security (ICCAIS), pages 1–2, Riyadh, Saudi Arabia, March 2020. IEEE.

[22] Waleed Al-Ghaith. Developing Lexicon-based Algorithms and Sentiment Lexicon for Sentiment Analysis of Saudi Dialect Tweets. International Journal of Advanced Computer Science and Applications, 10(11), 2019.

[23] Sarah Alhumoud, Tarfa Albuhairi, and Mawaheb Altuwaijri. Arabic Sentiment Analysis using WEKA a Hybrid Learning Approach:. In Proceedings of the 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management, pages 402–408, Lisbon, Portugal, 2015. SCITEPRESS - Science and and Technology Publications.

[24] Pansy Nandwani and Rupali Verma. A review on sentiment analysis and emotion detection from text. Social Network Analysis and Mining, 11(1):81, December 2021.

[25] MarkSchmidt,NicolasLeRoux,FrancisBach.MinimizingFiniteSumswitht heStochasticAverage Gradient.MathematicalProgramming,2017,162(1-2),pp.83-112.10.1007/s10107-016-1030-6

[26] Max Bramer, "Measuring the Performance of a Classifier," in Principles of Data Mining.: Springer, 2007, pp. 173-185.

[27] Wang, Xiaogang; Grimson, Eric (2007). "Spatial Latent Dirichlet Allocation" (PDF). Proceedings of Neural Information Processing Systems Conference (NIPS).

[28] Centers for Disease Control and Prevention. (2023, September 29). SARS-CoV-2 Variant Classifications and Definitions. Retrieved from https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-classifications.html

[29] World Health Organization. (2023, September 29). Coronavirus disease (COVID-19) pandemic - World Health Organization (WHO). Retrieved from https://www.who.int/europe/emergencies/situations/covid-19

[30] World Health Organization. (2020, April 27). Archived: WHO Timeline - COVID-19. Retrieved from https://www.who.int/news/item/27-04-2020-who-timeline---covid-19

[31] Kingdom of Saudi Arabia Ministry of Health. (2021, November 4). Saudi Ministry of Health Launches National Campaign to Raise Awareness About Influenza. Retrieved from https://www.moh.gov.sa/Ministry/MediaCenter/News/Pages/NEWS-2012-11-04-002.aspx

[32] Saudi Ministry of Health (@SaudiMOH). (2020, March 10). [Tweet in Arabic]. Retrieved from https://twitter.com/SaudiMOH/status/1234523092581523457?lang=ar

[33] Saudi Press Agency. (2021, November 3). (at: 30. September 2023) Retrieved from: https://stgcdn.spa.gov.sa/viewfullstory.php?lang=ar&newsid=2301159

[34] Saudi Press Agency. (2020, December 15). (at: 30. September 2023) Retrieved from: https://stgcdn.spa.gov.sa/viewfullstory.php?lang=ar&newsid=2168181

[35] Saudi Press Agency. (2021, July 23). (at: 30. September 2023) Retrieved from: https://stgcdn.spa.gov.sa/viewfullstory.php?lang=ar&newsid=2244988

[36] Saudi Press Agency. (2022, March 5). (at: 30. September 2023) Retrieved from: https://www.spa.gov.sa/2334814

[37] Jair Cervantes, Farid Garcia-Lamont, Lisbeth Rodríguez-Mazahua, Asdrubal Lopez, A comprehensive survey on support vector machine classification: Applications, challenges and trends, Neurocomputing, Volume 408, 2020, Pages 189-215, ISSN 0925-2312,

[38] Tessa Phillips, Waleed Abdulla, Developing a new ensemble approach with multi-class SVMs for Manuka honey quality classification, Applied Soft Computing, Volume 111, 2021, 107710, ISSN 1568-4946, https://doi.org/10.1016/j.asoc.2021.107710.

[39] Tweepy. (2023, July 18). Tweepy: A Python library for accessing the Twitter API. Retrieved from https://www.tweepy.org/

[40] Ekman P. (2007). Emotions revealed : recognizing faces and feelings to improve communication and emotional life (2nd ed.). Henry Holt.

[41] KNIME. (2023, March 8). KNIME Analytics Platform. Retrieved from https://www.knime.com/knime-analytics-platform

[42] Mohataher, M. (2023). Arabic stop words. GitHub. Retrieved from https://github.com/mohataher/arabic-stop-words.

[43] SaudiEAR . (2023, December 8) - Retrieved from (https://github.com/d7o-ae/saudiEAR/).