Heliyon 9 (2023) e19195

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Identifying drivers of COVID-19 vaccine sentiments for effective vaccination policy

Fahim Sufi [a,*], Musleh Alsulami [b]

[a] School of Public Health and Preventive Medicine, Monash University, 553 St. Kilda Rd., Melbourne, VIC, 3004, Australia
[b] Information Systems Department, Umm Al-Qura University (UQU), Makkah, Saudi Arabia

ARTICLE INFO

Keywords:
COVID-19 vaccine sentiment
Decision support system
Sentiment analysis
Entity detection
Regression analysis
Artificial intelligence

ABSTRACT

The COVID-19 pandemic has had far-reaching consequences globally, including a significant loss of lives, escalating unemployment rates, economic instability, deteriorating mental well-being, social conflicts, and even political discord. Vaccination, recognized as a pivotal measure in mitigating the adverse effects of COVID-19, has evoked a diverse range of sentiments worldwide. In particular, numerous users on social media platforms have expressed concerns regarding vaccine availability and potential side effects. Therefore, it is imperative for governmental authorities and senior health policy strategists to gain insights into the public's perspectives on vaccine mandates in order to effectively implement their vaccination initiatives. Despite the critical importance of comprehending the underlying factors influencing COVID-19 vaccine sentiment, the existing literature offers limited research studies on this subject matter. This paper presents an innovative methodology that harnesses Twitter data to extract sentiment pertaining to COVID-19 vaccination through the utilization of Artificial Intelligence techniques such as sentiment analysis, entity detection, linear regression, and logistic regression. The proposed methodology was applied and tested on live Twitter feeds containing COVID-19 vaccine-related tweets, spanning from February 14, 2021, to April 2, 2023. Notably, this approach successfully processed tweets in 45 languages originating from over 100 countries, enabling users to select from an extensive scenario space of approximately 3.55×10^{249} possible scenarios. By selecting specific scenarios, the proposed methodology effectively identified numerous determinants contributing to vaccine sentiment across iOS, Android, and Windows platforms. In comparison to previous studies documented in the existing literature, the presented solution emerges as the most robust in detecting the fundamental drivers of vaccine sentiment and demonstrates the vaccination sentiments over a substantially longer period exceeding 24 months.

1. Introduction

The coronavirus disease (COVID-19) pandemic has affected the lives of more than 766 million patients worldwide, out of which almost 7 million valuable lives have been prematurely lost (as of May 24, 2023) [1]. Other than the loss of precious lives, COVID-19 has caused global economic contraction, gross domestic product reduction, rising unemployment, higher levels of impoverishment, increasing income inequality, and other social conflicts, according to the World Bank [2]. Given this unforeseen global crisis,

* Corresponding author.
E-mail addresses: research@fahimsufi.com (F. Sufi), mhsulami@uqu.edu.sa (M. Alsulami).

https://doi.org/10.1016/j.heliyon.2023.e19195
Received 26 May 2023; Received in revised form 10 August 2023; Accepted 15 August 2023
Available online 19 August 2023
2405-8440/© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

governments in most countries have adopted mass vaccination drives against COVID-19 [1,2]. Hence, more than 10.5 billion vaccine doses have been administered successfully, as reported by the World Health Organization [1].

However, this abrupt vaccination drive has created confusion, agitation, and divisions across societies and cultures. There are anti-vaxxers who believe that the vaccine is a means for the government and corporate giants to control civilians [3,4]. The conflicts between anti-vaxxers and other communities have given rise to issues such as physical protests and riots, which have deep social impact [3–6]. Although a better understanding of sentiments on governments' vaccination mandates is essential, only a few recent studies have reported on this issue [7–11]. While the studies in both [7,8] were performed manually with online surveys, the studies in Refs. [9,10] were semiautomated and driven by data scientists using Twitter feeds. Data collected by Ref. [7] show that 18.85% of the surveyed population within the United Arab Emirates identified themselves as anti-vaxxers. More interestingly, more than 62.40% of the participants believed that there are more effective ways than vaccination to prevent disease [7]. In Ref. [8], researchers used statistical techniques, such as analysis of variance, regression, and k-nearest neighbor, on data gathered through questionnaires in order to identify the willingness and hesitancy regarding the vaccine drive in Qatar. In Ref. [9], sentiment analysis was performed on Persian tweets on keywords such as COVIran Barekat (the homegrown vaccine of Iran), Pfizer/BioNTech, AstraZeneca/Oxford, Moderna, and Sinopharm (imported vaccines) for 6 months. Further [10], used Twitter for analyzing sentiments towards AstraZeneca/Oxford, Pfizer/BioNTech and Moderna COVID-19 vaccines for 4 months. Nevertheless, these studies [7–10] on COVID-19 vaccine mandates suffers from the following shortcomings.

- They focused on a single region or country or city (e.g., UAE in Ref. [7], Qatar in Ref. [8], Iran in Ref. [9], Makkah and Madinah city of Saudi Arabia in Ref. [11]).
- They analyzed data in a very limited set of languages (e.g., English, Arabic or Persian).
- All these studies performed a longitudinal sentiment analysis for a short timeframe, such as 2 months [7] or 4 months [10].
- They were driven by specialists and required significant resources, such as the deployment of surveys, manual data cleansing, and analysis by data scientists.
- Most importantly, none of these studies identified the key drivers behind the vaccination sentiments.

To address the challenges and shortcomings in the literature as regards identifying the drivers of vaccination sentiment, in this study, we propose a methodology driven by artificial intelligence (AI) that will allow a strategic decision-maker to select a scenario (any country or group of countries, or any language or a group of languages, for any particular time interval). For the selected scenario, the proposed solution will instantly highlight all possible drivers along with the detailed correlation coefficients to signify the level of relationships. The proposed system works on Twitter feeds and is capable of comprehending tweets in more than 110 languages. As

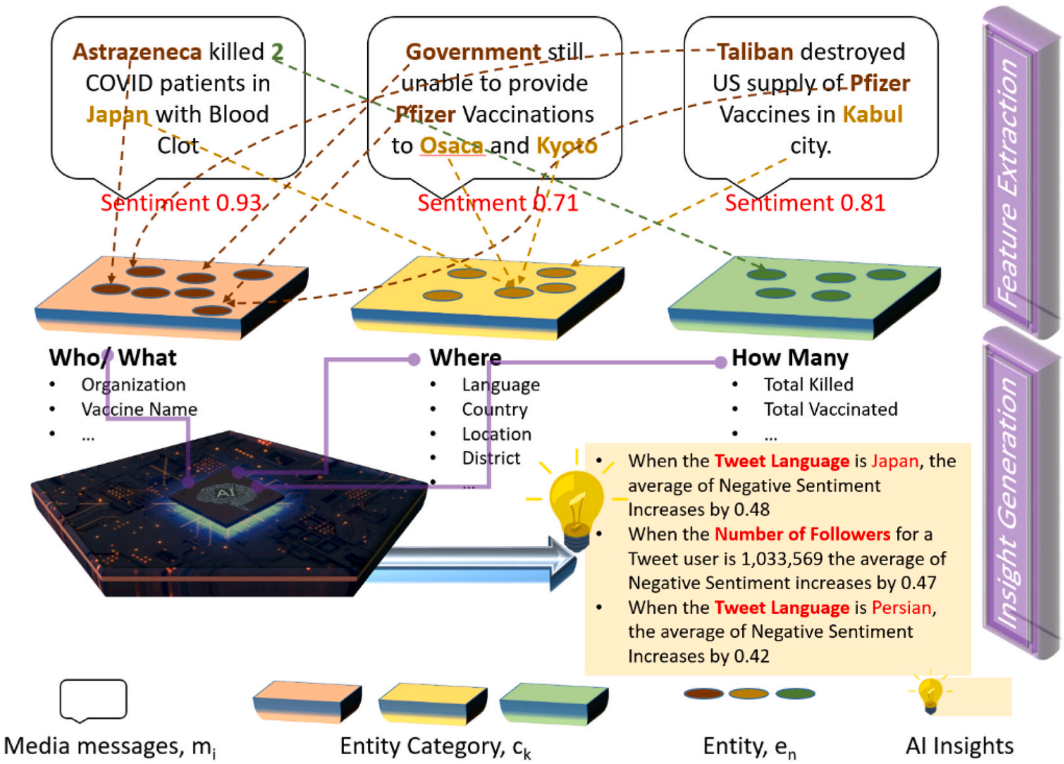

Fig. 1. Conceptual diagram of AI-based COVID-19 Vaccination Sentiment Analysis System.

Fig. 1 shows, by using AI-based techniques, such as sentiment analysis, named entity recognition (NER), and automated regressions (both linear, and logistic), the proposed solution identified the following drivers for a user defined scenario.

- When the tweet language is Japanese, the average negative sentiment grows by 0.48.
- When the number of followers of a Twitter user is more than 1,033,569, the average negative sentiment grows by 0.47
- When the tweet language is Persian, the average negative sentiment grows by 0.42.

The system was tested from February 14, 2021 to April 2, 2023, and during this 26-month period, the system analyzed tweets from more than 100 counties in 45 different languages. According to the existing reported literatures, this study is the longest longitudinal study that has investigated the key drivers behind vaccination sentiments. Moreover, the deployed solution was evaluated as extremely robust and supported an immense scenario-space of $3.55X10^{249}$. Last, the deployed solutions were tested on a diverse range of platforms, like iOS, Android, and Windows, with native apps. Hence, unlike the prior solutions in Refs. [7–11], the presented solution can be easily adopted by strategic decision-makers for making evidence-based decisions being assisted by the installed mobile apps.

2. Background

Sentiment recognition is the process of obtaining subjective knowledge from a text automatically with AI-based natural language processing (NLP). Using sentiment analysis, machines can learn whether a particular text contains a negative, a positive, or a neutral tone. Studies in the medical, biomedical, and health informatics fields have been using sentiment analysis for more than 20 years, as observed from Refs. [11–25]. In contrast, named entity extraction (NER) is the method of obtaining meaningful critical information, such as the person name, organization, product, location, and date/time information, from unstructured texts using NLP. Several studies in the medical, biomedical, and health informatics fields have used NER successfully to extract medically relevant information from unstructured blogs and social media messages, as noted from Refs. [15,20,21,26]. However, none of the existing studies has used NER to extract locations for COVID-19-related social media messages [11,14–19,27] .Table 1 provides a summary of the related literature, which shows that several studies have used sentiment analysis along with NER [15,20,21].

However, prior studies have not used sentiment analysis and NER as a preprocessor to extract feature attributes for further analysis by AI algorithms [15,20,21]. In a recent research [28], both sentiment analysis and NER were used to generate features automatically and then used the generated features to perform further AI-based analysis. In Ref. [28], Microsoft Power Bi [29], Microsoft Power Automate (or Microsoft Flow), and multiple services from Microsoft Azure were used. We tested the solution in Ref. [28] with live Twitter feeds from July 15, 2021, to August 10, 2021 that had the word "COVID" or "Corona." Our system recognized 24 different types of named-entities from these live Twitter feed and then clustered five different types of location-centric named-entities (i.e., Continent, Country Region, City, State, and Language) with aggregated sentiments in complete automation. Then, AI-based decomposition tree analysis was used for generating spatial intelligence from COVID-related Tweets [28,30,31]. The overall entity detection accuracy, including of all non-location entities, was 82.9% [28]. According to Ref. [28], the average precision and recall on the five location-centric named-entities were 0.901 and 0.967, respectively.

Sentiment recognition and NER techniques were also used to analyze global events [32,33], natural disaster detection [34], and to analyze political situational awareness [35].

In the present study, we report on identifying the drivers of COVID-19 vaccine sentiments by analyzing Twitter feeds using sentiment analysis, NER, and regression. The AI driven insights with linear and logistic regression can assist health care officials and strategic planners to make informed, timely, and evidence-based decisions on COVID-19 vaccinations.

Table 1
Literature review on the use of sentiment analysis and NER on medical research.

Reference	Language Supported	Sentiment Analysis	Entity Recognition	Location Detection	COVID-19/Corona Status	Automated Retrieval & Cleansing of Tweets
[11]	Arabic	Yes	No	No	Yes	No
[12]	English	Yes	No	No	No	No
[13]	English	Yes	No	No	No	No
[14]	English	Yes	No	No	Yes	No
[15]	English	Yes	Yes	No	Yes	No
[16]	English	Yes	No	No	Yes	No
[17]	English	Yes	No	No	Yes	No
[18]	English	Yes	No	No	Yes	No
[19]	Nepali	Yes	No	No	Yes	No
[26]	English	No	Yes	No	No	Yes
[20]	English	Yes	Yes	No	No	No
[21]	French	Yes	Yes	No	No	No
[22]	English	Yes	No	No	No	No
[23]	English	Yes	No	No	No	No
[27]	English	No	No	No	Yes	No
[24]	Arabic	Yes	No	No	No	No
[25]	English	Yes	No	No	No	No

3. Methodology

The proposed methodology uses a preprocessing stage to extract feature attributes from unstructured social media messages. As Fig. 2 shows, in the feature attribute extraction stage, AI-based language detection, translation, entity recognition, and sentiment analysis were used. Then, the results were saved in Microsoft SQL Server. When a user selects a scenario, appropriate data are fetched for that scenario in the next stage, termed "User-generated filtered records." In this stage, the appropriate data are retrieved to match the user-selected scenarios. Using these correct sets of records, regression analyses (both linear and logistic) are performed on the filtered records. Last, the results are presented to the user through different platforms, such as mobile, tablet, laptop, and desktop. In this section, details on the feature attribute extraction, user-generated filtered records, and the regression on filtered records are presented in detail.

A. Feature attribute extraction:

Feature extraction has many phases, such as filtering, language detection, translation, location detection, and sentiment analysis, as shown in Algorithm 1.

All Tweets are denoted by W, as seen from Eq. (1).

$$W = \{w_1, w_2, w_3, \ldots\}. \tag{1}$$

These messages are obtained with the Microsoft Power Automated process, whenever a tweet containing the keyword "COVID" or "Corona" is sensed. Since these tweets can be in any language, the language is detected with the *Detect_Language*(.) function as shown in Eq. (2).

$$Dectect\_Language(s_i) = \begin{cases} en, when\ s_i \in S\ and\ Language\ of\ s_i\ is\ English \\ ar, when\ s_i \in S\ and\ Language\ of\ s_i\ is\ Arabic \\ fr, when\ s_i \in S\ and\ Language\ of\ s_i\ is\ French \\ \ldots \\ da, when\ s_i \in S\ and\ Language\ of\ s_i\ is\ Danish \end{cases} \tag{2}$$

Then, the *Translate_English*(.) function translates the non-English tweets into English, as shown in Eq.. 3.

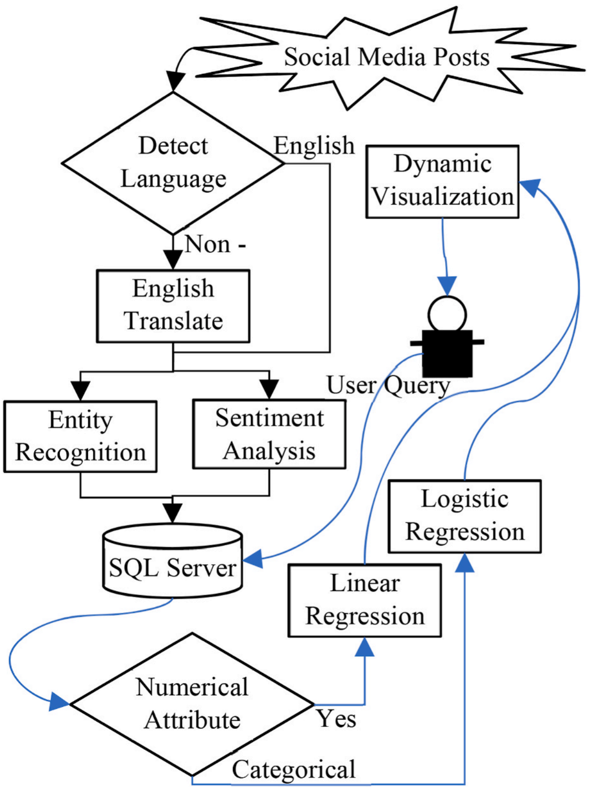

Fig. 2. Flow diagram of the proposed methodology.

$$M = \begin{cases} x : x = s_i, if\ l_i = en \\ x : x = Translate\_Engilish(s_i), if\ l_i <> en \end{cases} \tag{3}$$

The language detection function and the translation function both use the Application Programming Interface (API) of Microsoft Text Analytics and Cognitive Services [36,37]. At the end of this phase, all selected Twitter feeds are translated into a new set $M = \{m_1, m_2, m_3, \ldots, m_I\}$.

Then EntityDetection (.) detects entity $e_i^{m_i}$ and assigns the detected entities into the respective entity category c_k. In terms of implementation, NER API could be invoked from either Microsoft Power Automate or Microsoft Power Query Editor using DAX language. Code 1 is our implementation of calling NER from Microsoft Cognitive Services using DAX language (using Microsoft Power Query Editor). *DetectSentiment*(.) obtains the sentiment (i.e., value between 0 and 1) and the corresponding positive confidence score $s_i^{Positive}$ (value between 0 and 1), negative confidence score $s_i^{Negative}$ (value between 0 and 1), and neutral confidence score $s_i^{Neutral}$ (value between 0 and 1). Algorithm 1 completes the necessary preprocessing of the incoming real-time tweet feed with language detection, translation, NER, and sentiment analysis.

Algorithm 1
The Process of extracting features from Twitter feeds

Algorithm 1: *The Process of extracting features from Twitter feeds*

Input: *All the news descriptions,* $W = \{w_1, w_2, w_3, \ldots\}$

Output: *Complete feature attribute set,* $F_1, F_2, F_3, \ldots$ *or* $\{\{\{\{c_k, e_i^{m_i}\}\}, \{s_i, s_i^{Positive}, s_i^{Negative}, s_i^{Neutral}\}\}, \ldots\}$

For $\forall w_i \in W$

$m_i = Translate\_English(Detect\_Language(w_i))$

$\{\{c_k, e_i^{m_i}\}\} \leftarrow EntityDetection(m_i)$

$\{s_i, s_i^{Positive}, s_i^{Negative}, s_i^{Neutral}\} \leftarrow DetectSentiment(m_i)$

End Loop

Code 1
DAX code for invoking NER API of Text Analytics (i.e., Microsoft Cognitive Services)

```
(text) = > let
apikey = “##API-Keys-Goes-Here##”,
endpoint = “https://uaenorth.api.cognitive.microsoft.com/text/analytics/v2.1/entities”,
jsontext = Text.FromBinary(Json.FromValue(Text.Start(Text.Trim(text), 5000))),
jsonbody = “{ documents: [ { language: ““en”“, id: ““0”“, text: “ & jsontext & “ } ] }”,
bytesbody = Text.ToBinary(jsonbody),
headers = [#”Ocp-Apim-Subscription-Key” = apikey],
bytesresp = Web.Contents(endpoint, [Headers = headers, Content = bytesbody]),
jsonresp = Json.Document(bytesresp),
doc = jsonresp[documents]{0},
result = doc[entities],
#”Converted to Table” = Table.FromList(result, Splitter.SplitByNothing(.), null, null, ExtraValues.Error),
#”Expanded Column1” = Table.ExpandRecordColumn(#”Converted to Table”, “Column1”, {“name”, “matches”, “wikipediaLanguage”, “wikipediaId”,
“wikipediaUrl”, “bingId”, “type”}, {“name”, “matches”, “wikipediaLanguage”, “wikipediaId”, “wikipediaUrl”, “bingId”, “type”}),
#”Expanded matches” = Table.ExpandListColumn(#”Expanded Column1”, “matches”),
#”Expanded matches1” = Table.ExpandRecordColumn(#”Expanded matches”, “matches”, {“wikipediaScore”, “text”, “offset”, “length”,
“entityTypeScore”}, {“wikipediaScore”, “text”, “offset”, “length”, “entityTypeScore”})
in
#”Expanded matches1”
```

B. User-generated filtered records

Let all the COVID-19-related records be set A, the vaccine-related records be set B, Pfizer/BioNTech vaccine-related records be C, AstraZeneca/Covishield vaccine-related records be D, Moderna vaccine-related records be E, Sinopharm/Sinovac vaccine-related records be F, Sputnik vaccine-related records be G, and Covovax vaccine-related records be H. Then, Eq. (4) to Eq. (15) holds true.

$$C \subset B \subset A \tag{4}$$

The query for representing Pfizer/BioNTech is represented with Eq. (5):

$$C = \{x | (x\ contins\ ''COVID/CORONA'') \bigwedge (x\ contains\ ''VACCINE'') \bigwedge (x\ contains\ ''PFIZER'')\} \bigvee (contains\ ''BIONTECH'') \tag{5}$$

$$D \subset B \subset A \tag{6}$$

The query for representing AstraZeneca/Covishield is represented by

$$\mathrm{D} = \{x | (x\ contins\ ''COVID/CORONA'') \bigwedge (x\ contains\ ''VACCINE'') \bigwedge (x\ contains\ ''ASTRAZENECA'')\} \bigvee (contains\ ''COVISHIELD'') \tag{7}$$

$$E \subset B \subset A \tag{8}$$

The query for representing Moderna is represented by

$$\mathrm{E} = \{x | (x\ contins\ ''COVID/CORONA'') \bigwedge (x\ contains\ ''VACCINE'') \bigwedge (x\ contains\ ''MODERNA'')\} \tag{9}$$

$$F \subset B \subset A \tag{10}$$

The query for representing Sinopharm/Sinovac is represented by

$$\mathrm{F} = \{x | (x\ contins\ ''COVID/CORONA'') \bigwedge (x\ contains\ ''VACCINE'') \bigwedge (x\ contains\ ''SINOPHARM'')\} \bigvee (contains\ ''SINOVAC'') \tag{11}$$

$$G \subset B \subset A \tag{12}$$

The query for representing Sputnik is represented by

$$\mathrm{G} = \{x | (x\ contins\ ''COVID/CORONA'') \bigwedge (x\ contains\ ''VACCINE'') \bigwedge (x\ contains\ ''SPUTNIK'')\} \tag{13}$$

$$H \subset B \subset A \tag{14}$$

The query for representing Covovax is represented by

$$\mathrm{H} = \{x | (x\ contins\ ''COVID/CORONA'') \bigwedge (x\ contains\ ''VACCINE'') \bigwedge (x\ contains\ ''COVOVAX'')\} \tag{15}$$

Code 2 shows a sample SQL statement to fetch all the relevant records for "AstraZeneca" and "Covishield" vaccine from the Microsoft SQL Server Database using Eq. (7).

Code 2
SQL statement to fetch Astrazeneca/Covishield records

Code 2: *SQL statement to fetch Astrazeneca/Covishield records*

```
Select * FROM [dbo].[TweetDB] WHERE ([TweetText] like '%COVID%' or [TranslatedText] like '%COVID%' OR [TweetText] like '%CORONA%' or [TranslatedText] like '%CORONA%') AND ([TweetText] like '%VACCINE%' or [TranslatedText] like '%VACCINE%') AND ([TweetText] like '%ASTRAZENECA%' or [TranslatedText] like '%ASTRAZENECA%' OR [TweetText] like '%COVISHIELD%' or [TranslatedText] like '%COVISHIELD%')
```

Similarly, Eq. (4) to Eq. (15) could also be used from Microsoft Power Query environment to fetch relevant records using DAX language, as shown in Code 3.

Code 3 implements the logic demonstrated in Eq. (5) in statement 2. Then, in statement 3, it removes irrelevant columns. Since, for all the non-English tweets, Algorithm 1 translated the tweets in English, statement 4 of Code 3 creates a column called MasterTweet that holds original English tweets or translated tweets (in cases of non-English tweets). Last, in statements 5 and 6, the NER routine (as demonstrated in Code 1) is invoked and the necessary columns for the results of NER are fetched.

Code 3
DAX code for retrieving filtered tweets from SQL server

```
1. let
2. Source = Sql.Database("drsufiserver.database.windows.net", "SUFITWEETDB", [Query="Select * FROM [dbo].[Tweets] WHERE ([TweetText] like '%COVID%' or [TranslatedText] like '%COVID%' OR [TweetText] like '%CORONA%' or [TranslatedText] like '%CORONA%') AND ([TweetText] like '%VACCINE%' or [TranslatedText] like '%VACCINE%') AND ([TweetText] like '%PFIZER%' or [TranslatedText] like '%PFIZER%' OR [TweetText] like '%BIONTECH%' or [TranslatedText] like '%BIONTECH%')#(lf)"]),
3. #"Removed Columns" = Table.RemoveColumns(Source,{"UInterest"}),
4. #"Added Conditional Column" = Table.AddColumn(#"Removed Columns", "MasterTweet", each if [TweetLanguage] = "en" then [TweetText] else if [TweetLanguage] <> "en" then [TranslatedText] else null),
5. #"Invoked Custom Function" = Table.AddColumn (#"Added Conditional Column", "Entities", each EntityDetection ([MasterTweet])),
6. #"Expanded Entities" = Table.ExpandTableColumn(#"Invoked Custom Function", "Entities", {"name", "wikipediaScore", "text", "offset", "length", "entityTypeScore", "wikipediaLanguage", "wikipediaId", "wikipediaUrl", "bingId", "type"}, {"name", "wikipediaScore", "text", "offset", "length", "entityTypeScore", "wikipediaLanguage", "wikipediaId", "wikipediaUrl", "bingId", "type"}),
7. in
8. #"Expanded Entities"
```

C. Regression on filtered records

Regression analysis is utilized to recognize the most significant feature attributes in COVID-19 vaccine sentiments. Our deployment of regression analysis with Microsoft Power BI [29], automatically ranks the factors that drives COVID-19 vaccine sentiments, compares the significance of these factors, and displays them as key influencers. Two regression types (i.e., for both categorical and numeric metrics) are used, as shown in Fig. 5. Microsoft's ML. Net's SDCA regression [38] performed linear regression on the numerical features and deploying ML. Net's L-BFGS logistic regression [39,40] performed logistic regression on the categorical features. This process is demonstrated in recent studies on global landslide data [41,42], Bangladeshi Tornado Data [43,44], Australian Cyclones [45], global events [33,46], and other natural disasters [34,47]. Linear regression uses the equation in the form of Eq. (16).

$$y = b_0 + b_1 x_1 + \varepsilon \tag{16}$$

Logistic regression uses the equation in the form of Eq. (17).

$$\mathrm{Log}[y / y - 1] = b_0 + b_1 x_1 + b_2 x_2 + \ldots b_n x_n \tag{17}$$

The output of the logistic regression problem can only be between 0 and 1.

Logistic regression without a threshold is a regression [48]. However, the introduction of a threshold in the process transforms it into an efficient classifier. First, logistic or sigmoid function is used, as shown in Eq (18).

$$\sigma(t) = \frac{1}{1 + e^{-t}} \tag{18}$$

This translated real number to interval (0,1). Then, we use the hypothesis function in the form of Eq. (19).

$$h_\theta(x) = \sigma\left(\theta^T x\right) = \frac{1}{1 + e^{-\theta^T x}} \tag{19}$$

The classification decision is made on $y = 1$ when $h_\theta(x) \geq 0.5$ and $y = 0$ otherwise. The decision boundary is $\theta^T x = 0$. The cost function is represented with Eq. (20).

$$j(\theta) = \sum_{i=1}^{m} H\left(y^{(i)}, h_\theta\left(x^{(i)}\right)\right) \tag{20}$$

where H(p,q) is the cross-entropy of distribution q relative to distribution p. It is shown with Eq. (21).

$$H(p, q) = -\sum_i p_i \log q_i \tag{21}$$

In this case, $y^{(i)} \in \{0,1\}$ so $p_1 = 1$ and $p_2 = 0$; Hence, the relationship is represented with Eq. (22).

$$H\left(y^{(i)}, h_\theta\left(x^{(i)}\right)\right) = -y^{(i)} \log h_\theta\left(x^{(i)}\right) - \left(1 - y^{(i)}\right) \log\left(1 - h_\theta\left(x^{(i)}\right)\right) \tag{22}$$

Similarly, to the selection of the quadratic cost function in linear regression, the selection of this cost function is demonstrated with Eq. (23).

$$\mathrm{grad}\ \mathrm{J}(\theta) = \frac{\partial \mathrm{J}(\theta)}{\partial \theta} = \begin{bmatrix} \frac{\partial}{\partial \theta_0} \mathrm{J}(\theta) \\ \frac{\partial}{\partial \theta_1} \mathrm{J}(\theta) \\ \cdot \\ \cdot \\ \cdot \\ \frac{\partial}{\partial \theta_n} \mathrm{J}(\theta) \end{bmatrix} = \mathrm{X}^{\mathrm{T}}\left(\mathrm{h}_\theta(\mathrm{X}) - \mathrm{y}\right) \tag{23}$$

Thus, the gradient descent for logistic regression is represented with Eq. (24).

$$\theta(k+1) = \theta(k) - s\ \mathrm{grad}\ J(\theta) \tag{24}$$

By default, logistic regression provides binary classification. To build a classifier with L classes, L binary classifiers (i.e., one vs. rest) are used.

D Obtaining Ai-Driven Insights

The heat map generation process is used in this study to detect the location-centric high-level and low-level COVID-19-related tweets. The Getis-Ord Gi* statistic is the method used to identify statistically meaningful spatial clusters at the local scale [49]. We have used this technique in our recent study on disaster monitoring from social media [34]. The final outcome of the Heat-Map Algorithm is the important hot- and cool-spot areas, characterized by Eq. (25), Eq. (26), and Eq. (27).

$$G_i^* = \frac{\sum_{j=1}^{n} w_{i,j}x_j - \overline{X}\sum_{j=1}^{n} w_{i,j}}{S\sqrt{\frac{\left[n\sum_{j=1}^{n} w_{i,j}^2 - \left(\sum_{j=1}^{n} w_{i,j}\right)^2\right]}{n-1}}} \tag{25}$$

where x_j is the attribute value for feature j, $w_{i,j}$ is the spatial weight between feature i and j, n is equal to the total number of features, and:

$$\overline{X} = \frac{\sum_{j=1}^{n} x_j}{n} \tag{26}$$

$$S = \sqrt{\frac{\sum_{j=1}^{n} x_j^2}{n} - (\overline{X})^2} \tag{27}$$

Algorithm 2 shows how location information is first extracted from the Twitter feeds by only segregating the location entities of relevant location entity categories. Then, only location entities with greater than 70% confidence level are extracted and used for obtaining longitude and latitude information with the GeoDecode(.) function. Last, GenerateHeatmap(.) implements Eq. (25) to Eq. (27) for generating heatmaps. Notably, GenerateHeatmap(.) takes both longitude and latitude information along with the sentiment of the extracted location entity for generating heat maps.

Algorithm 2
Location detection from Twitter feeds

Algorithm 2: *Location detection from Twitter feeds*

Input*: All the Tweet Texts,* $M = \{m_1, m_2, m_3, \ldots\}$

Output*: Complete feature attribute set,* $F_1, F_2, F_3, \ldots$ *or* $\{\{\{\{c_k, e_i^{m_i}\}\}, \{s_i, s_i^{Positive}, s_i^{Negative}, s_i^{Neutral}\}, g_i\}, \ldots\}$

For $\forall m_i \in M$

  $\{\{c_k, e_i^{m_i}, d_i^{m_i}\}\} \leftarrow EntityDetection(m_i)$

  $if\,(c_k = \text{City})||(c_k = \text{Country Region})||(c_k = \text{State})||(c_k = \text{Language})||(c_k = \text{Continent})||(c_k = \text{StreetAddress})$

    $if\,(d_i^{m_i} > 0.7)$

      $L \leftarrow \{e_i^{m_i}\}$

End Loop

For $\forall l_i \in L$

  $GenerateHeatmap(GeoDecode(l_i), s_i)$

End Loop

4. Results

This study analyzed randomly selected records of 147,966 COVID-19-related tweets gathered from February 14, 2021, to April 2, 2023. As Table 2 shows, in the total records, there were 13,919 records containing the keyword Vaccine. There were 1258 records containing the keyword combination Pfizer and BioNTech (as represented previously with Equation (5)). There were 122 records on AstraZeneca and Covishield (represented with Equation (7)). Table 3 provides further details on these selected records in terms of the number of users, retweet counts, follower counts, friend count, locations, and languages.

This paper reports the outcomes of the longest longitudinal study for monitoring COVID-19 vaccine sentiments, that is, from February 2021 to April 2023. As Fig. 3 shows, the average negative sentiment worldwide about the Pfizer vaccine peaked during May 2021 and January 2022. The average negative sentiments about the AstraZeneca vaccine peaked during February 2021 (see Fig. 4). As Fig. 5 shows, the negative sentiments (average) were highest during November 2021. Conversely, the public negative perceptions about the Chinese-made Sinovac were at its peak during August 2021 and then slowly declined in later months (Fig. 6). Last, the Russian-made Sputnik vaccine accumulated the most negative sentiments in October 2021, and then, May 2022 (see Fig. 7). However, the negative perception towards Sputnik is gradually increasing since November 2021.

Fig. 8 shows a scenario for which Pfizer vaccine-related tweets in 40 different languages were analyzed between February 14, 2021, and April 2, 2023. The top five tweet languages for this scenario were English (72.51%), Arabic (6.28%), Turkish (4.5%), Persian/Farsi (3.99%), and French (2.37%). The Twitter users had about 611 million followers and 11 million friends, and these tweets had been retweeted about 3 million times. The follower count, friend count, and retweet count represent the impact and significance of these tweets. For these 1258 tweets, our implementation of NER (with Code 1) detected 10,803 different entities classified into seven different entity categories. These entity categories were date/time, location, organization, person, quantity, URL, and others. Fig. 8

represents the following drivers of negative sentiments towards the Pfizer vaccine for this particular scenario.

1) When the tweet language is ja (i.e., Japanese), the average confidence of negative sentiment grows by 0.48.
2) When the follower count is more than 1,033,569, the average confidence of negative sentiment grows by 0.47.
3) When the tweet language is fa (i.e., Persian), the average confidence of negative sentiment grows by 0.42.
4) When the tweet language is it (i.e., Italian), the average confidence of negative sentiment grows by 0.4.
5) When the tweet language is tr (i.e., Turkish), the average confidence of negative sentiment grows by 0.27.
6) … And 13 more drivers behind negative sentiments on Pfizer

As Fig. 8 shows, using this proposed methodology, a strategic decision-maker can obtain critical insights immediately upon selecting a particular scenario.

Fig. 9 represents another scenario where all the conditions are almost identical to those of the previous scenario, except that the user had selected a specific date range, May 20, 2021, to September 6, 2021. Immediately, the following drivers of the negative sentiment on Pfizer are calculated by AI.

1) When the tweet language is pl (i.e., Polish), the average confidence of negative sentiment grows by 0.62.
2) When the tweet language is ja (i.e., Japanese), the average confidence of negative sentiment grows by 0.58.
3) When the friend count is between 56 and 62, the average confidence of negative sentiment grows by 0.43.
4) When the tweet language is tr (i.e., Turkish), the average confidence of negative sentiment grows by 0.43.
5) When the tweet language is fi (i.e., Finnish), the average confidence of negative sentiment grows by 0.41.
6) When the tweet language is it (i.e., Italian), the average confidence of negative sentiment grows by 0.4.
7) … And 15 more drivers behind negative sentiments on Pfizer

Fig. 10 shows that our regression-based AI solution can also identify specific clusters; using the NLP technique, our solution explains these findings to the user in plain English. As Fig. 10 shows, within segment 1 the average negative sentiment is 0.75 (i.e., a very high negative perception toward Pfizer vaccination), which is 0.37 units higher than the overall average sentiment. This segment 1, representing a cluster with significantly higher negative sentiment confidence, is characterized by the following.

1) The follower count is less than or equal to 729.
2) The retweet count is less than or equal to 0 or greater than 373.
3) Tweets contain hashtags for regional defense ministries.

These characteristics are often demonstrated by misinformation spread by anti-vaxxers.

Figs. 8–10 represented cases with categorical variables, and hence, logistic regressions were performed; Fig. 11 presents a scenario when the relationship is depicted with linear regression. For this case, the user selected tweets in the Persian language between June 2, 2021, and October 27, 2021. From the selected cases, the following relationships were discovered.

1) When the follower count is 317 or less, the average confidence of negative sentiment grows by 0.69 (discovered by logistic regression).
2) When the friend count reduces by 719.44, the average confidence of negative sentiment grows by 0.07 (discovered by linear regression).

5. Dicussion

In this section, first, we test the robustness of the deployed and tested system. Then, we briefly discuss the deployment of the solution into multiple platforms and services. Last, we present the limitations of this study.

Table 2
Number of Records Per Keyword Combination

Serial #	Keyword Combinations	No. Of Records
1	Records with COVID + CORONA	147966
2	Vaccine	13919
3	Pfizer	1194
4	Pfizer + BioNTech	1258
5	AstraZeneca	99
6	AstraZeneca + Covishield	122
7	Moderna	279
8	Sinopharm + Sinovac	32
9	Sputnik	38
10	Covovax	1

Table 3
Number of Records Per Keyword Combination.

Keyword Combinations	No. Of Records	No. Of Unique Users	Retweet Counts	Follower Counts	Friend Count	Number of Unique Locations	Number of Unique Language
Vaccine	13919	11327	15392619	1021042068	18101557	3723	43
Pfizer + BioNTech	1258	1157	21120731	628127447	17342581	523	23
AstraZeneca + Covishield	122	103	57758	15905267	142767	58	15
Moderna	279	257	438686	4564816	356395	124	18
Sinopharm + Sinovac	32	28	1245	4207764	18263	18	8
Sputnik	38	36	6584	451415	59481	25	10
Covovax	1	1	1	18036	424	1	1

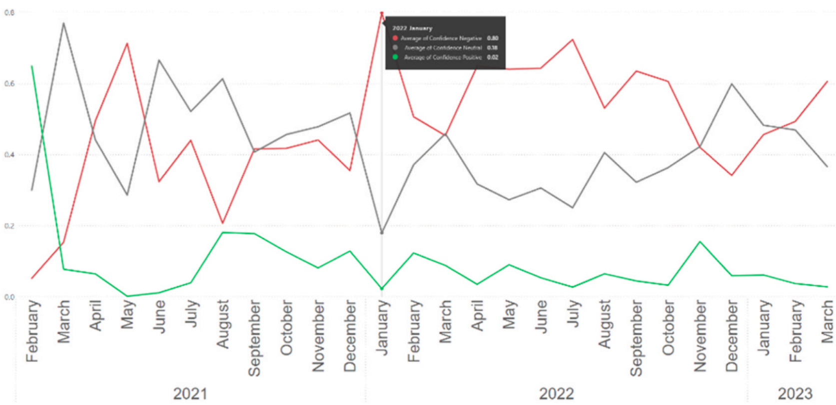

Fig. 3. Longitudinal study showing the sentiment about Pfizer vaccine for more than 24 months.

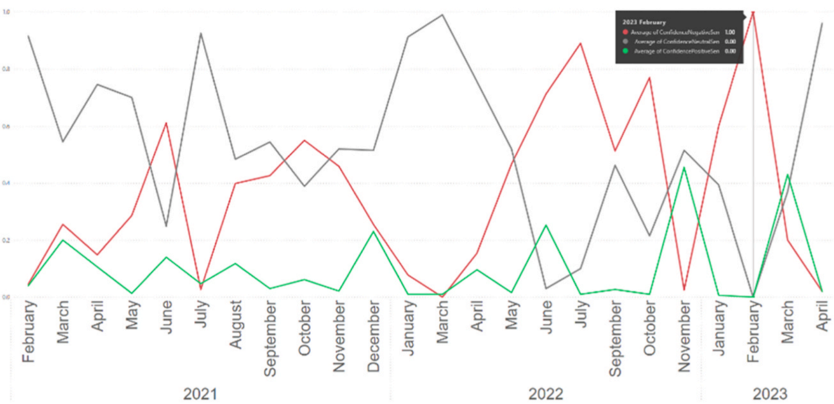

Fig. 4. Longitudinal study showing the sentiment about AstraZeneca vaccine for 24 months.

A. Evaluation of robustness:

To evaluate its robustness, we calculated the number of scenarios supported by the deployed solution. As Figs. 8–11 show, the deployed solution allowed users to select any combination of languages, dates, and entity categories. Hence, to calculate the permutations or combinations of these filter options (i.e., dates, languages, entity categories). Consider a hypothetical situation. If a selector allowed three options, such as Pfizer, AstraZeneca, and Sinovac, then it could support the following seven filter options.

1. {*Pfizer*}
2. {*AstraZeneca*}
3. {*Sinovac*}
4. {*Pfizer*, *AstraZeneca*}

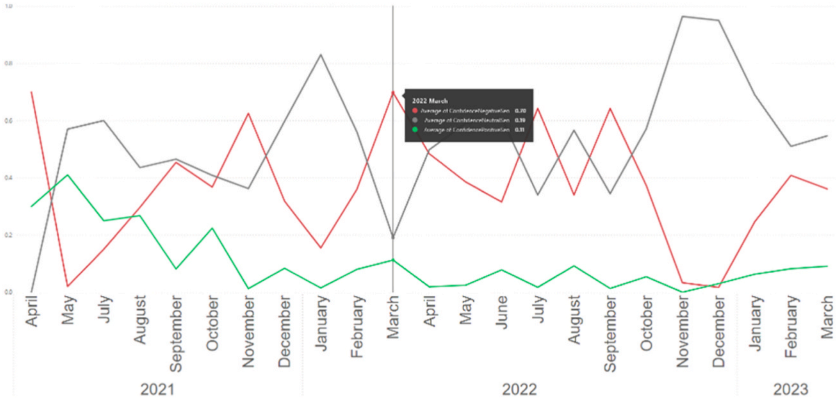

Fig. 5. Longitudinal study showing the sentiment about Moderna vaccine for 24 months.

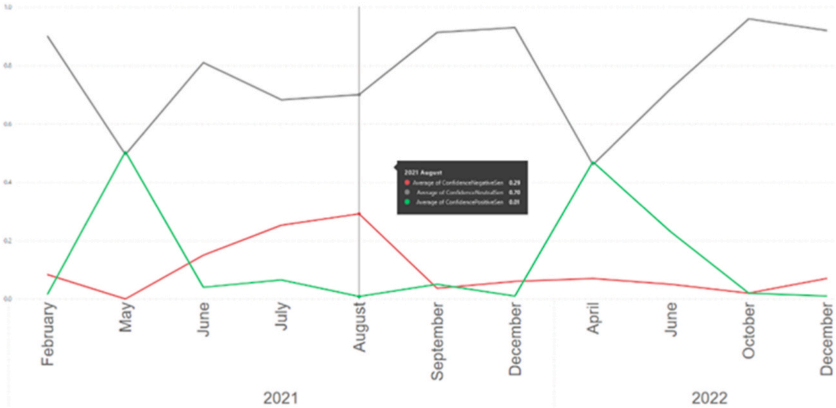

Fig. 6. Longitudinal study showing the sentiment about Sinovac vaccine.

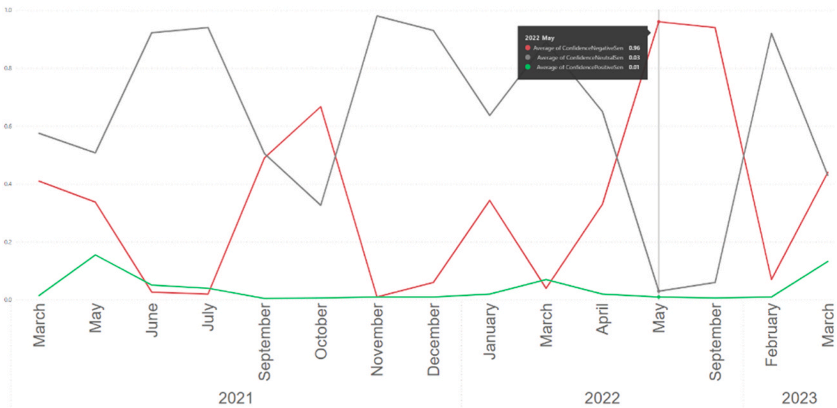

Fig. 7. Longitudinal study showing the sentiment about Sputnik vaccine.

5. {*AstraZeneca*, *Sinovac*}
6. {*Pfizer*, *Sinovac*}
7. {*Pfizer*, *AstraZeneca*, *Sinovac*}

Therefore, for this hypothetical selector, there could be seven possible filter settings as represented by $(2^T - 1)$, which is the formula to calculate the power set of selector option minus 1 (i.e., $P(T) - 1$). That is, 1 is deducted since the power set also includes an empty set and the selection of an empty set is not a supported option for our visualization, as depicted in Figs. 8–11. Hence, the total

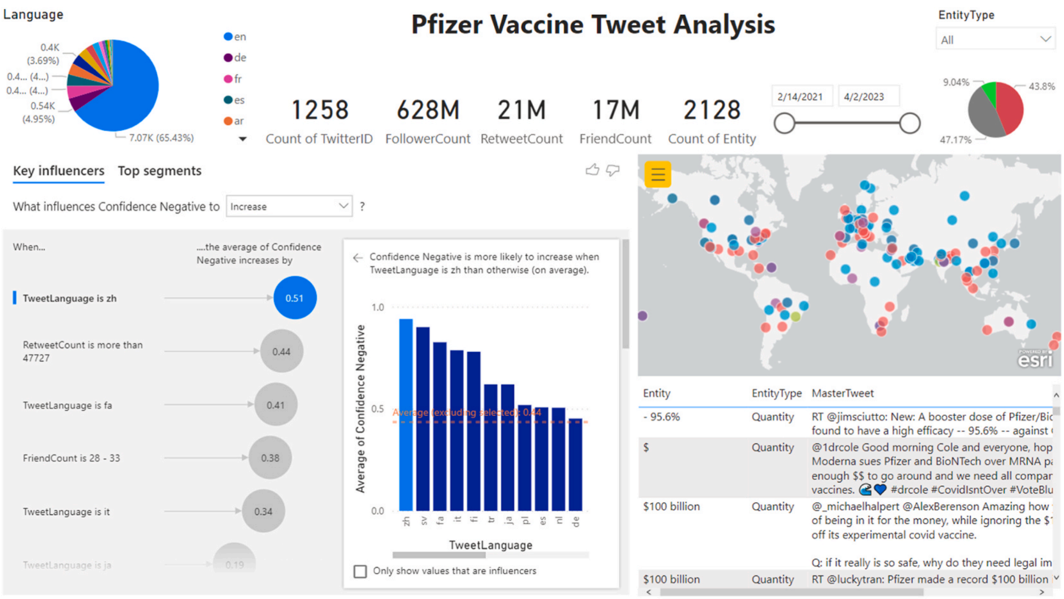

Fig. 8. Visualization of the drivers behind negative sentiment of Pfizer vaccine (Case 1).

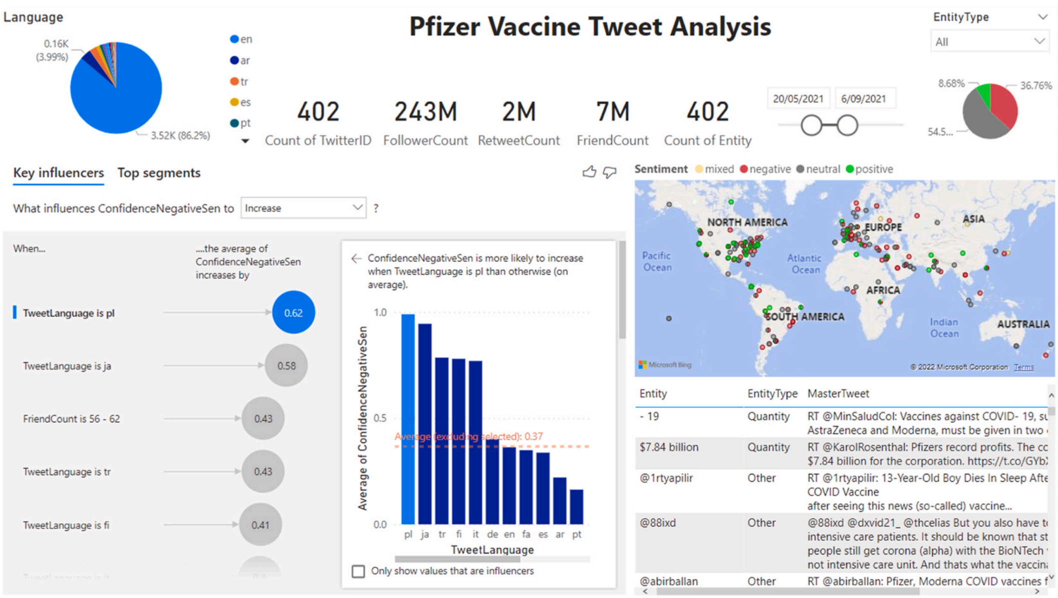

Fig. 9. Visualization of the drivers behind negative sentiment of Pfizer vaccine (Case 2).

number of possible scenarios can be calculated as

$$|S| = \left(2^{777} - 1\right) X \left(2^{45} - 1\right) X \left(2^{7} - 1\right) = 3.55X10^{249} \tag{28}$$

Equation (28) was calculated on the basis of 777 days between February 14, 2021 and April 2, 2023, 45 supported languages, and seven different entity categories. The purpose of this section is not to produce an exhaustive list of insights from the COVID-19 vaccination data, but to demonstrate the ability of the designed AI solution to produce insights on any scenario out of the $3.55X10^{249}$ possible scenarios (as shown in Equation (28)). It should be highlighted that no previous solution for vaccine sentiment monitoring (i.e. [7–10]) has demonstrated as high robustness as our presented solution.

B. Evaluation of performance:

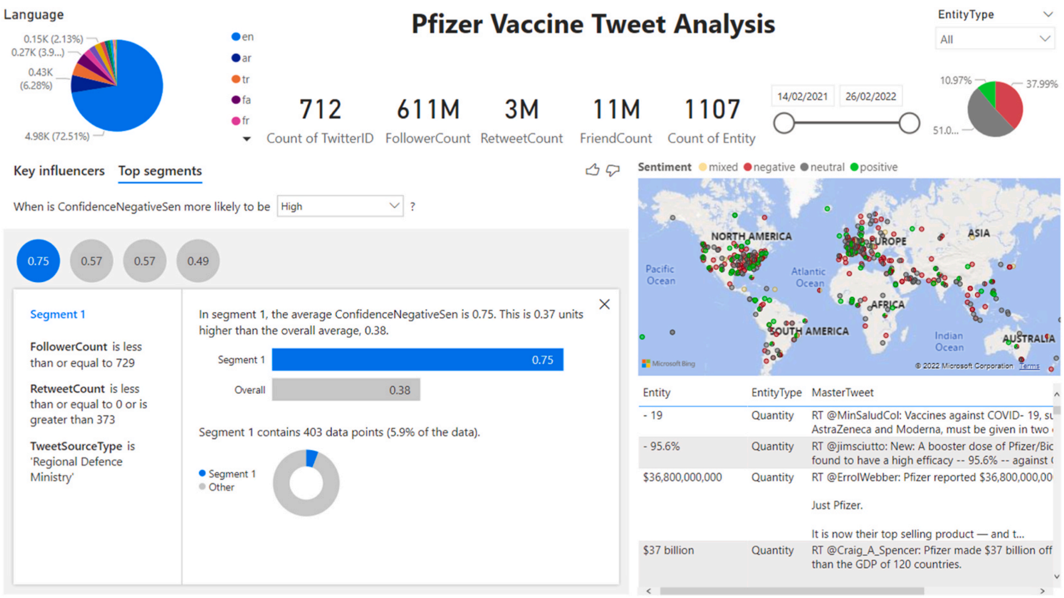

Fig. 10. Visualization of the drivers behind negative sentiment of Pfizer vaccine (Case 3).

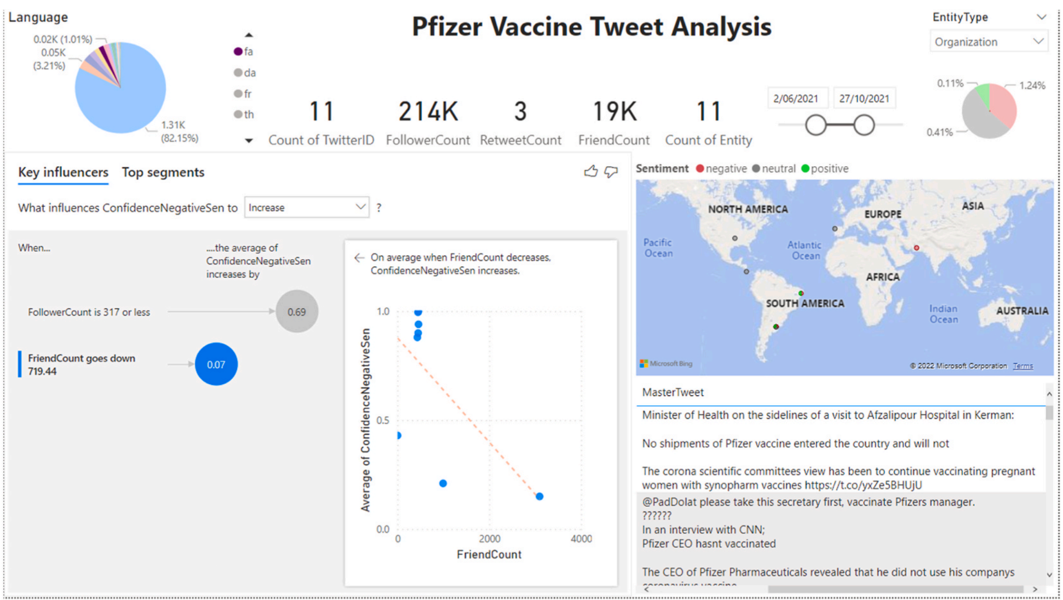

Fig. 11. Visualization of the drivers behind negative sentiment of Pfizer vaccine (Case 4).

Evaluation was performed with 1258 Tweets corresponding to Pfizer and BioNTech category of vaccines. Out of these 1258 Tweets, 7058 Tweets were in English, 535 Tweets were in German, 492 in French, 437 in Spanish, 436 in Arabic, 399 in Farsi/Persian, 353 in Turkish, 285 in Japanese and 15 more languages. While the language detection part could successfully detect all the 23 different languages for Pfizer and BioNTech category with 100% Accuracy, the translation had some errors. Manual evaluation of all the non-English language to English language requires assistance from expert linguists of all the 22 languages for this case. However, we only evaluated the appropriateness of translation from a single language pair, which is 436 Arabic Tweets. For Arabic translation, the accuracy was about 94%. While accuracy varies with the language pair, according to Microsoft [50], overall accuracy of machine translation for common language pair is around 80%–90% (as reported in Ref. [50]). Next, we performed performance evaluation of Sentiment Analysis for the 1258 Tweets corresponding to Pfizer + BioNTech category as seen from Table 4. The overall, precision, recall, and F1-Score were found to be 0.98, 0.96, and 0.97. Moreover, we also performed evaluation of NER in terms of precision, recall, and F1-Score as shown in Table 5. The 10,803 entities detected from the 1258 Pfizer (as well as BioNTech) Tweets with NER is shown in Table 5. The average precision, recall, and F1-Score were found to be 96%. It should be mentioned, using the same method of NER (i.e., Microsoft Cognitive Service) a previous studies in Refs. [32,46] obtained a higher precision, recall, and F1-Score of 99%. In Ref. [34], NER based location detection obtained overall precision of 0.93, recall of 0.88, and F1-score of 90% (with 97% accuracy) from live Tweeter Feed. Within that same study (i.e. [46]), Mean Absolute Percentage Error (MAPE) was found to be 0.255 on an average for linear regression and Area Under Curve (AUC) was found to be 0.895 for logistic algorithm. Within [41], Microsoft Power BI's automated regression achieved a slightly higher performance (i.e., AUC = 0.911, MAPE = 0.255). Since this study used automated regression provided by Microsoft Power BI (i.e., similar to previous studies in Refs. [41,46]) on Twitter feed, AUC was 0.9 on an average.

The presented methodology was deployed in Microsoft Platform, Microsoft SQL Server, and we utilized several APIs of Microsoft Azure Cognitive Services, as in our prior studies in Refs. [32,33,35,41,42]. Hence, our selected architecture allowed us to deploy the solution seamlessly onto multiple platforms, such as mobiles, tablets, desktops, and laptops through native apps. We tested our solution on the iOS, Android, and Windows platform. Fig. 12 shows the deployment in the Android environment (Android 12) on a Samsung Galaxy Note 10 Lite mobile. Fig. 13 demonstrates a user running our deployed solution in Apple iOS version 15 on an iPad 9th Generation. Figs. 8–10 represent our deployed solution in the Microsoft Windows environment. Since strategic decision-makers may not always be in their office when they need to make quick decisions related to vaccine sentiments, they can use their mobile phones (be it Android, Windows, or event iOS) and use the proposed solution to make evidence-based decisions.

Even though our platform supports big data with millions of records, as shown in Refs. [32,33,35,41,42], for this study, we performed regression analysis on approximately 147,966 randomly selected tweets related to COVID-19/coronavirus (out of which the number of vaccine-related tweets were only 13,919). Moreover, we only utilized regression algorithms to discover the drivers behind negative sentiments on vaccines from tweet texts and did not analyze sentiments using images or videos posted by social media users.

E. User notes

Microsoft Power BI Desktop was used within this study and this tool can be downloaded from Ref. [29]. Once Microsoft Power BI is downloaded and installed, the sources files (i.e., the. pbix file) could be downloaded from Ref. [31]. Clicking the. pbix file, allows the source code along with the entire solution to be loaded under Microsoft Power BI Desktop environment.

6. Conclusion

This study has proposed a new methodology for discovering the drivers behind vaccine sentiments using AI-based techniques, namely, language detection, translation, entity detection, sentiment analysis, linear regression, and logistic regression. A typical user of this system would be strategic decision-makers, policymakers, diplomats, and general users who wish to know about the vaccine sentiments in any region or country (since this system is capable of generating insights for more than 110 countries). The presented system enables critical analysis of location-centric vaccine sentiments providing advice on policy implementation in order to mitigate risks associated with COVID-19 vaccinations. In short, the following are the core contributions of this presented study.

- Our deployed solution executed the presented algorithms from February 14, 2021 to April 2, 2023 and thus provided results on vaccine sentiments for the longest period compared with existing literature [7–10].
- Our deployed solution is extremely robust and supports an enormous scenario space of $\mathbf{3.55\textit{X}10^{249}}$. Unlike the previous studies in Refs. [7–10], users of the presented system can select any scenarios out of this extremely large scenario space, following which the presented solution automatically executes a regression analysis immediately to discover all the drivers behind vaccine sentiments. Understanding the drivers behind vaccine sentiments are specifically important as COVID-19 is becoming more predictable and manageable (i.e., Endemic [51]). As shown in this study, the drivers behind vaccine sentiments were friend count, follower could, presence of certain hashtags (like, regional defense ministries), and most importantly Tweet languages (specifically for languages like Polish, Japanese, Persian, Turkish, Finnish, Italian). Microsoft Power BI's retrogression discovered correlation of these driving factors with negative sentiment.
- The presented solution is the only solution that can be seamlessly deployed into multiple platforms, such as iOS, Android, and Windows.

In future, we will endeavor to use multiple algorithms, including greedy correlation-based feature selection, expectation maximization clustering, k-mean clustering, and convolution neural networks, and other machine learning and statistical algorithms as shown in previous research in Refs. [32,41,52–54].

Table 4
Performance of Sentiment Analysis (Evaluated with 1258 Tweets corresponding to Pfizer + BioNTech).

	TP	FP	FN	Precision	Recall	F1-Score
Negative	488	15	33	0.97	0.94	0.95
Positive	104	2	4	0.98	0.96	0.97
Neutral	589	9	14	0.98	0.98	0.98
				0.98	0.96	0.97

Table 5
Performance of NER (From 1258 Tweets of Pfizer + BioNTech Categories that had 10,803 Entities).

C. Support of multiple platforms:

	TP	FP	FN	Precision	Recall	F1-Score
Date/Time	1534	32	45	0.98	0.97	0.98
Location	1833	56	29	0.97	0.98	0.98
Organization	1213	87	93	0.93	0.93	0.93
Person	2110	27	47	0.99	0.98	0.98
Quantity	1099	62	39	0.95	0.97	0.96
URL	1041	74	57	0.93	0.95	0.94
Other	1235	29	61	0.98	0.95	0.96

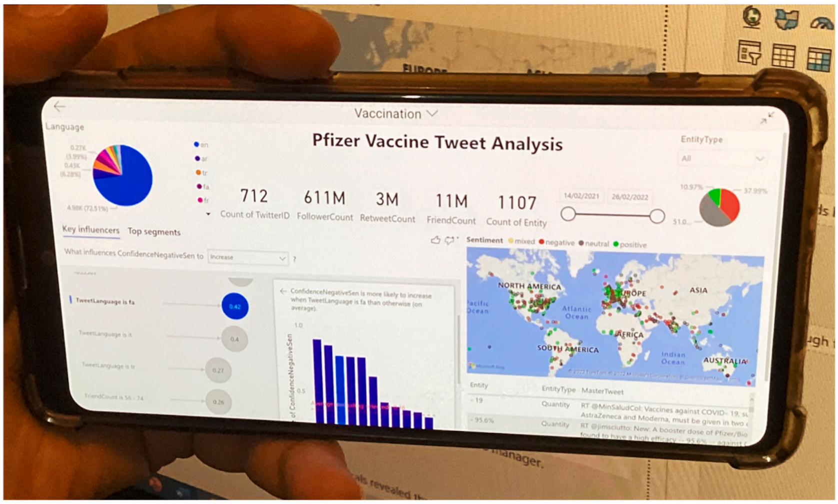

Fig. 12. Deployed solution in Android App on Samsung Galaxy Note 10 Lite.

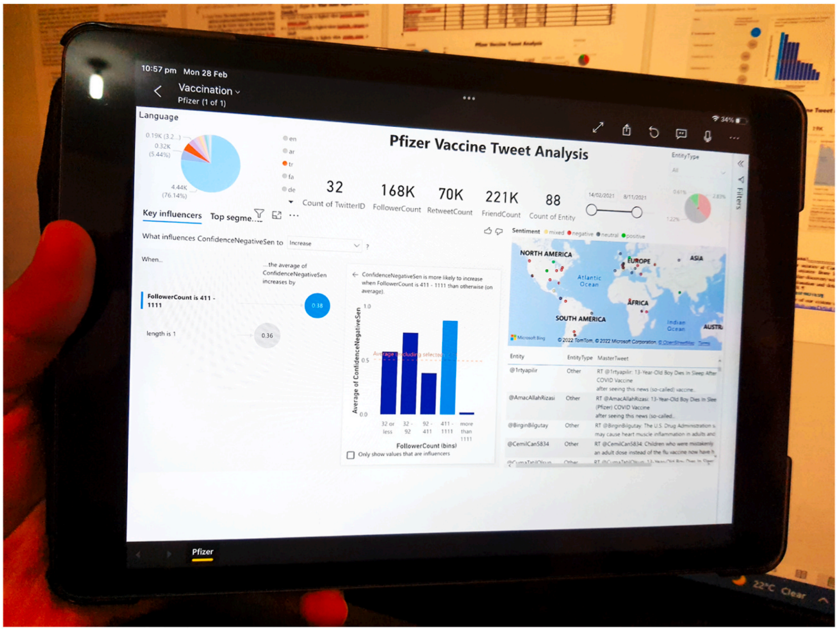

Fig. 13. Deployed solution in iOS App on Apple iPad 9th Generation.

D. Limitation of work:

Author contribution statement

Fahim Sufi: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper. Musleh Alsulami: Contributed reagents, materials, analysis tools or data.

Data availability statement

Data will be made available on request.

Additional information

No additional information is available for this paper.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to thank the Deanship of Scientific Research at Umm Al-Qura University for supporting this work by Grant Code: (23UQU4290525DSR003). The authors would also like to thank COEUS Institute, Kennebunk, Maine, US, where the first author works as the Chief Technology Officer.

References

[1] WHO, "WHO Coronavirus, COVID-19) Dashboard [Online]. Available:, 2022 https://covid19.who.int/. (Accessed 1 March 2022).
[2] A. Tandon, T. Roubal, L. McDonald, P. Cowley, T. Palu, V. de Oliveira Cruz, P. Eozenou, J. Cain, H.S. Teo, M. Schmidt, E. Pambudi, I. Postolovska, D. Evans, C. Kurowski, Economic Impact of COVID-19 : Implications for Health Financing in Asia and Pacific, 2021 [Online]. Available: https://openknowledge.worldbank.org/handle/10986/34572. (Accessed 1 March 2022).
[3] N.F. Johnson, N. Velásquez, N.J. Restrepo, R. Leahy, N. Gabriel, S.E. Oud, M. Zheng, P. Manrique, S. Wuchty, Y. Lupu, The online competition between pro- and anti-vaccination views, Nature 582 (2020) 230–233.
[4] F. Sufi, I. Razzak, I. Khalil, Tracking anti-vax social movement using AI based social media monitoring, IEEE Transactions on Technology and Society 3 (4) (2022) 290–299.
[5] Reuters, Melbourne Police Fire Pepper Balls, Pellets to Break up COVID-19 Protest, 21 Sep, 2021 [Online]. Available: https://www.reuters.com/world/asia-pacific/melbourne-construction-sites-shut-down-after-anti-vaccine-mandate-protest-2021-09-20/. (Accessed 1 January 2022).
[6] Evening Standard, Police Suffer Injuries during Anti-vaxxer Protest in Parliament Square, 18 Dec, 2021 [Online]. Available: https://www.standard.co.uk/news/uk/police-injuries-anti-vaxxer-protest-parliament-square-b972758.html. (Accessed 1 January 2022).
[7] R. Muqattash, I. Niankara, R.I. Traoret, Survey data for COVID-19 vaccine preference analysis in the United Arab Emirates, Data Brief 33 (2020), 106446.
[8] M. Hafizh, Y. Badri, S. Mahmud, A. Hafez, P. Choe, COVID-19 vaccine willingness and hesitancy among residents in Qatar: a quantitative analysis based on machine learning, J. Hum. Behav. Soc. Environ. (2021) 1–24, https://doi.org/10.1080/10911359.2021.1973642.
[9] Z.B. Nezhad, M.A. Deihimi, Twitter sentiment analysis from Iran about COVID 19 vaccine, Diabetes Metabol. Syndr.: Clin. Res. Rev. 16 (1) (2022), 102367.
[10] R. Marcec, R. Likic, Using twitter for sentiment analysis towards AstraZeneca/oxford, pfizer/BioNTech and Moderna COVID-19 vaccines, Postgrad. Med. (2021), https://doi.org/10.1136/postgradmedj-2021-140685.
[11] M.K. Shambour, A. Gutub, The impact of corona disease on the feelings of twitter users in the two holy cities during the hajj season of 1442 AH, in: Hajj Forum 2022 - the 21st Scientific Hajj Research Forum, 2022. Madinah , Saudi Arabia.
[12] B. Pang, L. Lee, S. Vaithyanathan, Thumbs up?: sentiment classification using machine learning techniques, in: Conf. Empirical Methods Natural Lang, Process., 2002.
[13] P.D. Turney, Thumbs up or thumbs down?: semantic orientation applied, in: *40th* Annu. Meeting, 2002.
[14] U. Naseem, I. Razzak, M. Khushi, P.W. Eklund, J. Kim, COVIDSenti: a large-scale benchmark twitter, IEEE Transactions on computational social systems 8 (4) (2021) 1003–1015.
[15] S. Boon-Itt, Y. Skunkan, Public perception of the COVID-19 pandemic on twitter: sentiment analysis and topic modeling study, JMIR Public Health Surveill 6 (4) (2020).
[16] L. Nemes, A. Kiss, Social media sentiment analysis based on COVID-19, Journal of Information and Telecommunication 5 (1) (2020) 1–15.
[17] K.H. Manguri, R.N. Ramadhan, P.R.M. Amin, Twitter sentiment analysis on worldwide COVID-19 outbreaks, Kurdistan Journal of Applied Research (KJAR) 5 (3) (2020) 54–63.
[18] L. Li, Q. Zhang, X. Wang, J. Zhang, Characterizing the propagation of situational information in social media during COVID-19 epidemic: a case study on weibo, IEEE Transactions on computational social systems 7 (2) (2020) 556–562.
[19] C. Sitaula, A. Basnet, A. Mainali, T.B. Shahi, Deep Learning-Based Methods for Sentiment Analysis on Nepali COVID-19-Related Tweets, Computational Intelligence and Neuroscience, 2021, 2158184.
[20] D. Cameron, G.A. Smith, R. Daniulaityte, A.P. Sheth, D. Dave, L. Chen, G. Anand, R. Carlson, K.Z. Watkins, R. Falck, PREDOSE: a semantic web platform for drug abuse epidemiology using social media, J. Biomed. Inf. 46 (6) (2013).
[21] X. Chen, C. Faviez, S. Schuck, A. Lillo-Le-Louët, N. Texier, B. Dahamna, C. Huot, P. Foulquié, S. Pereira, V. Leroux, P. Karapetiantz, A. Guenegou-Arnoux, S. Katsahian, C. Bousquet, A. Burgun, Mining patients' narratives in social media for pharmacovigilance: adverse effects and misuse of methylphenidate, Front. Pharmacol. 9 (541) (2018).
[22] B. Chan, A. Lopez, U. Sarkar, The Canary in the Coal Mine Tweets: Social Media Reveals Public Perceptions of Non-medical Use of Opioids, PLOS One, 2015.
[23] E.C. McNaughton, R.A. Black, M.G. Zulueta, S.H. Budman, S.F. Butler, Measuring online endorsement of prescription opioids abuse: an integrative methodology, Pharmacoepidemiol. Drug Saf. 21 (10) (2012) 1081–1092.
[24] N. Al-Twairesh, H. Al-Negheimish, Surface and Deep Features Ensemble for Sentiment Analysis of Arabic Tweets, IEEE Access, 2019, pp. 84122–84131.
[25] G. Vashisht, Y.N. Sinha, Sentimental study of CAA by location-based tweets, Int. J. Inf. Technol. 13 (2021) 1555–1567.

[26] E. Batbaatar, K.H. Ryu, Ontology-based healthcare named entity recognition from twitter messages using a recurrent neural network approach, Int. J. Environ. Res. Publ. Health 16 (3628) (2019).

[27] I. Arpaci, S. Alshehabi, M. Al-Emran, M. Khasawneh, I. Mahariq, T. Abdeljawad, A.E. Hassanien, Analysis of twitter data using evolutionary clustering during the COVID-19 pandemic, Computers, Materials & Continua CMC 65 (1) (2020) 193–203.

[28] F. Sufi, M. Alsulami, AI-Based automated extraction of location-oriented COVID-19 sentiments, Comput. Mater. Continua (CMC) Accepted (Press) (2022).

[29] B.I. Microsoft Power, Download Power BI Desktop, 2021 [Online]. Available: https://app.powerbi.com/.

[30] B.I. Microsoft Power, Entity Detection and Sentiment Analysis on COVID-19 Situation, 17 8, 2021 [Online]. Available: https://app.powerbi.com/view?r=eyJrIjoiZGQzNDhhYjctYjZkZi00ZDllLWExZDgtZTJjMjk1NTQ4ZTc4IiwidCI6IjBkMWI4YmRlLWZmYzEtNGY1Yy05NjAwLTJhNzUzZGFjYmEwNSJ9.

[31] GitHub, GitHub Repository of COVID-19 Entities with Sentiment Analysis [Online]. Available:, 2021 https://github.com/DrSufi/COVID_Sentiment_Entity.

[32] F. Sufi, M. Alsulami, Automated multidimensional analysis of global events with entity detection, sentiment analysis and anomaly detection, IEEE Access 9 (2021).

[33] F. Sufi, AI-GlobalEvents: a Software for analyzing, identifying and explaining global events with Artificial Intelligence, Software Impacts 11 (2022), 100218.

[34] F. Sufi, I. Khalil, Automated disaster monitoring from social media Posts using AI based location intelligence and sentiment analysis, IEEE Transactions on Computational Social Systems (2022) 1–11, https://doi.org/10.1109/TCSS.2022.3157142. Accepted (in Press).

[35] F. Sufi, M. Alsulami, A novel method of generating geospatial intelligence from social media Posts of political leaders, Information 13 (3) (2022) 120.

[36] Microsoft Documentation, "https://docs.microsoft.com/en-us/azure/cognitive-services/text-analytics/how-tos/text-analytics-how-to-language-detection," 07 02 2021. [Online]. Available: https://docs.microsoft.com/en-us/azure/cognitive-services/text-analytics/how-tos/text-analytics-how-to-language-detection. [Accessed 20 August 2021].

[37] Microsoft Documentation, Microsoft text Analytics API: languate detection [Online]. Available: https://docs.microsoft.com/en-us/azure/cognitive-services/text-analytics/language-support?tabs=language-detection. (Accessed 10 August 2021).

[38] H.-F. Yu, C.-J. Hsieh, K.-W. Chang, C.-J. Lin, Large linear classification when data cannot fit in memory, in: KDD '10: Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2010.

[39] H. Matthies and G. Strang, "The solution of non linear finite element equations," Int. J. Numer. Methods Eng., vol. 14, no. 11, p. 1613–1626..

[40] J. Nocedal, Updating quasi-Newton matrices with limited storage, Math. Comput. 35 (151) (1980) 773–782.

[41] F.K. Sufi, M. Alsulami, Knowledge discovery of global landslides using automated machine learning algorithms, IEEE Access 9 (2021) 131400–131419.

[42] F. Sufi, AI-Landslide: software for acquiring hidden insights from global landslide data using Artificial Intelligence, Software Impacts 10 (2001), 100177.

[43] F. Sufi, E. Alam, M. Alsulami, A new decision support system for analyzing factors of Tornado related deaths in Bangladesh, Sustainability 14 (10) (2022) 6303.

[44] F.K. Sufi, Ai-Tornado, An AI-based Software for analyzing Tornadoes from disaster event dataset, Software Impacts 11 (2022) 1–5.

[45] F. Sufi, E. Alam, M. Alsulami, Automated analysis of Australian tropical Cyclones with regression, clustering and convolutional neural network, Sustainability 14 (16) (2022) 9830.

[46] F. Sufi, Identifying the drivers of negative news with sentiment, entity and regression analysis, International Journal of Information Management Data Insights 2 (1) (2022), 100074.

[47] F. Sufi, Ai-SocialDisaster, An AI-based software for identifying and analyzing natural disasters from social media, Software Imapcts 13 (2022), 100319.

[48] B. Kisacanin, Short introduction to deep learning and its math - Part 1, Mathematical Reflections 2 (2020) 1–10.

[49] J.K. Ord, A. Getis, Local spatial autocorrelation statistics: distributional issues and an application, Geogr. Anal. 27 (4) (1995) 286–306.

[50] Aorinka, What Automatic Translation Service Is the Most Accurate [Online]. Available:, 2023 https://www.linguise.com/blog/guide/what-automatic-translation-service-is-the-most-accurate/. (Accessed 3 July 2023).

[51] C. Klobucista, M. Ferragamo, When will COVID-19 become endemic? [Online]. Available: https://www.cfr.org/in-brief/when-will-covid-19-become-endemic, 2023. (Accessed 3 July 2023).

[52] P. Roy, S. Saumya, J. Singh, S. Banerjee, A. Gutub, Analysis of Community Question-answering Issues via Machine Learning and Deep Learning: State-of-the-art Review, CAAI Transactions on Intelligence Technology, 2022, pp. 1–23.

[53] S.J. Alghamdi, Determining the Mix Design Method for Normal Strength Concrete Using Machine Learning, vol. 14, Journal of Umm Al-Qura University for Engineering and Architecture, 2023, pp. 95–104.

[54] M.A. Abu-Hashem, A. Gutub, O. Salem, M.K. Shambour, Q. Shambour, M. Shehab, A. Izzat, M.J. Alrawashdeh, Discrepancies of Remote Techno-Tolerance Due to COVID-19 Pandemic within Arab Middle-East Countries, Journal of Umm Al-Qura University for Engineering and Architecture, 2023, https://doi.org/10.1007/s43995-023-00026-0. Early Access:.