Bulletin of Electrical Engineering and Informatics
Vol. 12, No. 3, June 2023, pp. 1648~1656
ISSN: 2302-9285, DOI: 10.11591/eei.v12i3.4278

Predicting COVID-19 vaccinators based on machine learning and sentiment analysis

Hadab Khalid Obayes[1], Khaldoon Hasan Alhussayni[2], Saba Mohammed Hussain[3]
[1]College of Education for Humanities Studies, University of Babylon, Babylon, Iraq
[2]Computer Center, University of Babylon, Babylon, Iraq
[3]Department of Information Networks, University of Babylon, Babylon, Iraq

Article Info

Article history:

Received Jun 17, 2022
Revised Oct 18, 2022
Accepted Nov 2, 2022

Keywords:

COVID-19
Machine learning
Predicting
Sentiment analysis
Vaccine

ABSTRACT

In the past two years, the world witnessed the spread of the coronavirus (COVID-19) pandemic that disrupted the entire world, the only solution to this epidemic was health isolation, and with it everything stopped. When announcing the availability of a vaccine, the world was divided over the effectiveness and harms of this vaccine. This article provides an analysis of vaccinators and analysis of people's opinions of the vaccine's efficacy and whether negative or positive. Then a model is built to predict the future numbers of vaccinators and a model that predicts the number of negative opinions or tweets. The model consists of three stages: first, converting data sets into a synchronized time series, that is, the same place and time for vaccination and tweets. The second stage is building a prediction model and the third stage was descripting analysis of the prediction results. The autoregressive integrated moving averages (ARIMA) method was used after decomposing the components of ARIMA and choosing the optimal model, the best results obtained from seasonal ARIMA (SARIMA) for both predictions, the last stage is the descriptive analysis of the results and linking them together to obtain an analysis describing the change in the number of vaccinators and the number of negative tweets.

Corresponding Author:

Hadab Khalid Obayes
College of Education for Humanities Studies, University of Babylon
Babylon, Iraq
Email: Hedhab@uobabylon.edu.iq

1. INTRODUCTION

The world is going through a transition; there is a change in how day-to-day activities are handled during the continuing coronavirus (COVID-19) epidemic-whether it's e-learning or the way people socialize, engage, utilize, or store things [1]. At this time, it is essential to take the necessary precautions to protect oneself, including washing one's hands often, wearing a mask when going into close quarters, and avoiding unnecessary physical contact. However, the effects of these interventions are limited to preventing the further spread of the COVID-19 rather than eliminating it. At this point, vaccination started to play an essential role as the only method that had any chance of controlling and ultimately eliminating the COVID-19 [2]. Extensive testing was carried out using the first mRNA vaccines to be made available for purchase. Over 40,000 individuals participated in the Pfizer vaccination experiment, while over 30,000 people participated in the vaccine trial conducted by Moderna [3]. The manufacture of various vaccinations is a significant challenge; nevertheless, the startling lack of people's desire and motivation to get vaccinated is even more disturbing and of considerable worry to health specialists interested in determining the reasons behind this phenomenon. Since its inception, the vaccination procedure has been met with ambivalent reactions from the general public; even within our own families, we have been subjected to disagreements or questions over this topic [4]. People use social media platforms to share their thoughts, opinions, and reactions to better manage and respond to extreme crises.

This is essential for social media platforms to play in severe problems because it allows people to collaborate on crisis management and response [5]. The COVID-19 has been one of the moving topics on Twitter since January 2020 and continues to be investigated to this day [6]. In this way, users of Twitter can easily get the sentiments of their tweets out to the people [7]. Predicting the numbers of people vaccinated against the COVID-19 and analyzing their feelings about the types of vaccines were among the encouraging topics for researchers. The tweet represents the user's opinion, and if sentiment analysis is relied upon to know the user's opinion, it can be based on sentiment analysis and consider it an influential element in people's conviction of the quality of the product [8]. The article's Alam *et al.* [4] propose a technique for analyzing tweets.

The natural language processing (NLP) technology known as valence aware dictionary (VADER) was used to investigate individuals' feelings towards every kind of vaccination. Reasoning based on sentiment VADER. It was helpful to conceptualize the entire situation by dividing the polarity of the received feelings into three groups: positive, negative, and neutral. The findings showed that 33.96% of the replies were favourable, 17.55% were negative, and 48.49% were neutral [9]. They also included a timeline analysis of tweets in this poll because respondents' emotions changed over time. When using long short-term memory (LSTM) and recurrent neural networks (RNN), including bidirectional (BiLSTMs), to test the performance of the prediction models, LSTMs obtain an accuracy of 90.59%. The BiLSTMs have a success rate of 90.83%. Many other performance measures such as accuracy, F1 score, confusion matrix, and so on were also employed to verify the model and the findings correctly. This work contributes to the objective of eradicating COVID-19 all over the world by enhancing the public's awareness of how COVID-19 may be prevented by vaccination. This study [6] predicted the popularity of tweets by analyzing public opinion and sentiment at different stages of the COVID-19 pandemic, from disease outbreak to vaccine distribution. Five sets of content features were extracted and applied to supervised machine learning algorithms, including topic analysis, topics and term frequency inverse document frequency (TF-IDF) vectorizers, TF-IDF vectorizer wordbags (BOW), document embedding, document embedding, and TF-IDF vectorizers. You posted a tweet to generate it.

According to the analysis, tweets with high emotional strength are more popular than tweets with information about the COVID-19 pandemic. Based on two statistical models, as well as the deep learning (DL) model, the author of the paper [10] does an analysis and makes a forecast about the daily number of confirmed cases of COVID-19. DNN with long-term memory, autoregressive integrated moving averages (ARIMA), generalized autoregressive conditional heteroskedasticity (GARCH), and GARCH stacked on top of each other. With the use of autocorrelation and partial autocorrelation functions, as well as an exhaustive search for DL model hyperparameters like the number of LSTM cells and cell blocks, the order of the statistical model may be identified. The experiment utilizes 10 data sets. Conduct research on how factors such as data size and inoculation affect performance. The numerical findings demonstrate that the performance depends on the data that were utilized and the data that were initially used. It is also shown that LST MDNN can produce more accurate forecasts compared to the two statistical models. According to the experiments' findings, LSTM DNN can improve up to 88.54% (86.63%) and 90.15% (87.74%), respectively.

According to Ardabili *et al.* [11] analyzed and evaluated the capabilities of several machine learning models to forecast the spread of COVID-19 in the United States of America, China, Iran, Germany, and Italy. The models' findings using an adaptive neural fuzzy inference system and multilayer perceptron (MLP) showed some encouraging signs. Research has shown that machine learning models are better than other approaches to simulating the COVID-19 epidemic. A further recommendation in the study was to predict death rates in order to anticipate the demand for critical care beds. At the end of the day, she advocated combining machine learning with susceptible-exposed-infectious removed (SEIR) models to enhance the accuracy and timeliness of typical epidemiologic models. These models take into account susceptible individuals who have been exposed to infectious individuals who have been removed. It was found that multiple supervised machine learning approaches were employed to mimic COVID-19 infection in Mexico in the recently published work [12]. Analyses were performed using a variety of models, including SVM, logistic regression, decision trees, Naive Bayes, and artificial neural networks. The study of correlation coefficients was used to examine how the dataset's characteristics are related to one another. The findings showed that the accuracy of the decision tree model was the greatest possible, coming in at 94.99%. Naive Bayes' specificity was 94.30%, while the SVM's sensitivity was 93.34%. There were seven alternative regression models employed by [13] to forecast the number of infected Egyptians (exponential, logit, quadratic, third, fourth, fifth, and sixth-degree). They trained the models using data from the official database, which was available from February 15 to June 15 of the subsequent year. After 15 days, one month, and one month, these models precisely anticipated the formation of COVID-19 and its final magnitude and longevity. These models were shown to be most accurate in predicting future events for 15 days afterward. On the other hand, the fourth-degree model's predictive power was shown over one month. Using the logit growth regression model, the pandemic determination would reach its zenith on June 22, 2020, and that it would end on September 8, 2020. In addition, it was anticipated that there would be a total of 166,760 cases of the pandemic. However, the presented findings could not be trusted entirely because of the existing social and environmental (climatic) circumstances.

2. THEORETICAL BASIS FOR THE POPOSED MODEL

2.1. Sentiment analysis

The word "sentimental analysis" refers to various activities, including deducing findings, defining assessments, organizing subjectivity, cataloguing assumptions, and identifying spam. SA intends to research people's presumptions, attitudes, mentalities, conclusions, sentiments, and so on concerning things, people, problems, associations, administrations, and so on [14]. In the field of sentiment analysis, several methodologies, such as lexical-based methods and supervised machine learning, may assist us in determining how people are feeling about something. Learning by machine necessitates the use of training data, which may be challenging to get. In addition, the training process takes a significant amount of time and is computationally demanding regarding the needs placed on the CPU and memory [15]. A set of linguistic properties that are commonly categorized as positive or negative according to their semantic orientation constitutes what is known as a sentiment lexicon. The majority of research in sentiment analysis makes use of preexisting lexicons that were developed by human labor. This is because establishing a lexicon is a difficult task. LIWC 1, GI 2, and Hu-Liu04 3 are the lexicons used most of the time [16].

2.2. TextBlob

Developed in Python, Textblob is a tool for manipulating large amounts of text. For the purpose of conducting NLP operations, it provides a standard application programming interface (API) [9]. It is identical to a string written in Python [17]. It uses a sentiment lexicon and a sentiment analysis engine called pattern.en. Pattern.en analyzes the text based on the English adjectives included in it, and WordNet is utilized to do this. TextBlob will produce a tuple of the form (polarity, subjectivity) whenever it does sentiment analysis on a text. The polarity value will be a float that falls in range [-1,1] [16]. One of the advantages of using TextBlob is that its strings are pretty similar to those of Python. TextBlob's operation will become easier to use. In addition to tokenization and noun phrase extraction, Textblob contains capabilities such as sentiment analysis, point-of-sales tagging, language translation and detection, n-grams, spelling correction, and interactivity with WordNet [17].

2.3. Time series data and prediction

Records of observations made on a specific topic throughout numerous periods make up what is known as time-series data. Collecting observable data with a uniform distribution may be defined as a time series. These data are also collected at consistent intervals. Data analysis methods are used in time series analysis to describe the data and derive meaning and usefulness from statistical information. A model is used in time series forecasting, which predicts future values based on values that have been observed in the past. Concerning the procedures part of this process, the forecast is based only on historical data. It operates under the presumption that the exact causes that affect the past and present will also impact the future [18]. If it is possible to deduce the values of a time series's future observations from its previous observations, then the time series is unavoidable (deterministic). If the importance of the time series may partly determine the future of a time series in the past, then the time series in question can be described as stochastic or random. Successful linear approaches include ARMA and ARIMA, both of which are linear models; nevertheless, the predictive capacity of linear models is constrained by the linear behavior of the underlying data [19].

2.4. ARIMA model

As a shorthand for the Box-Jenkins model, the ARIMA acronym stands for (p, d, q). parameter p denotes the order of autoregression, parameter d denotes difference, and parameter q denotes moving average. " For the ARIMA model, the letters "AR," "MA," and "I" represent autoregressive, moving average, and integration, respectively [20]. ARIMA models for stationary time series have the following mathematical representations: [3]. Autoregressive model of order p or AR(p) model:

$$y_t = \delta + \emptyset_1 y_{t-1} + \emptyset_2 y_{t-2} + \cdots + \emptyset_p y_{t-p} + \varepsilon_t \tag{1}$$

Moving-average model or order q or MA(q):

$$y_t = \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t \tag{2}$$

Autoregressive moving average model of order p and q or ARMA (p,q):

$$y_t = \delta + \emptyset_1 y_{t-1} + \cdots + \emptyset_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} \quad (3)$$

where ∅ is the parameter for autoregression and θ is the parameter for moving average. yt represents the actual value at time t, while? is the constant. The random disturbance term is εt considered white noise, with a mean of 0 and a variance shared by all times [21]. In recent years, the ARIMA model has emerged as one of the most used strategies for predicting epidemics. This approach is well suited for the short-term prediction of infectious illnesses, and the accuracy of its predictions has been extensively acknowledged; it may also provide practical support for disease prevention and policy development [22].

Metrics and statistical models used:

In this study, the following are the primary metrics and statistical models that were employed (where Xi represents the actual data, for instance, I and Pi represent the prediction [23]. The mean absolute error (MAE) measures the average number of mistakes for a group of forecasts, even though it does not consider the direction in which the errors are going. The average of the absolute differences the sample exhibited between the estimates and the actual observation, considering that each deviation is of equivalent weight [24].

$$MAE = \frac{\sum_{i=1}^{n} |X_i - P_i|}{n} \quad (4)$$

Mean absolute percentage error (MAPE) quantifies the precision of a forecasting system. This accuracy is expressed as a percentage, which may be computed as the average absolute% inaccuracy for each period minus the actual values divided by the actual values [25].

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{X_i - P_i}{X_i} \right| \quad (5)$$

3. METHOD

The availability of data on COVID 19 in all its variants, including the number of infections, the number of recoveries, the number of deaths, the number vaccinated, and the public's opinion of vaccination, has prompted many researchers to use these data. This article analyzes the data on the number of vaccinees and develops a model to predict the future number of vaccinees. It also analyzes people's opinions about the vaccine by dividing the tweets into positive, negative, and normal. ARIMA was used to predict the number of future vaccinees and was used to predict the number of negative tweets about a vaccine. Figure 1 shows the proposed model. After forming the data series for each of the vaccinees dataset and the tweets with the same date, i.e. day and month, the TextBlob library is used to analyze and classify the tweets (positive, negative and normal).

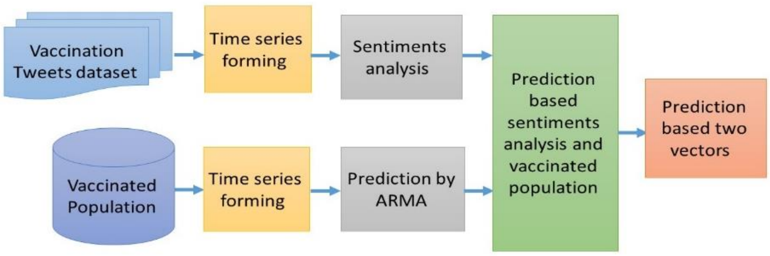

Figure 1. Method of the proposed model

3.1. Prediction for vaccinations

For the number of vaccinations, ARIMA is used for prediction. Figure 2 shows the total number of vaccinations Pfizer in the top 10 countries form 12/12/2020 to 23/11/2021. Figure 3 shows the decomposition of ARIMA to the vaccinations. From the previous analysis, we may infer that there is an "upward trend" in overall vaccination rates. Therefore, this time series is 'non-stationary,' and based on the seasonal component, we may conclude that the model is 'additive' since the seasonal component remains constant (i.e., it does not become multiplied) across time. Best model: ARIMA (2,2,1) (0,0,0) [0]. Automated model selection determines seasonal ARIMA (SARIMA) (2,2,1) to be the optimal model based on AIC. Prob(Q)=0.82>0.05. We should not reject the null hypothesis that residuals are uncorrelated since they are not correlated. Prob(JB)=0.00<0.05. The null hypothesis that residuals are regularly distributed is rejected. The residuals are thus regularly distributed. The model's residuals appear to exhibit correlation. ARIMA is the sole model that rejects the Jarque-Bera hypothesis (Prob (JB) 0.05). Therefore, the residuals for this model have a normal distribution.

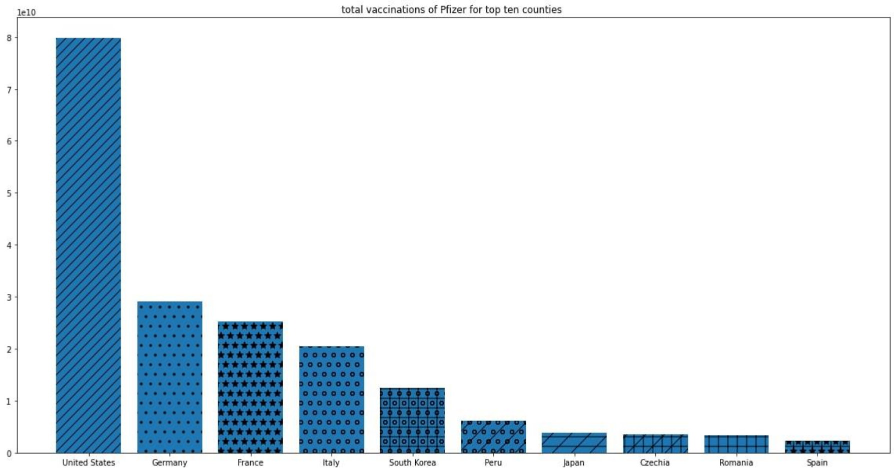

Figure 2. The total vaccinations in top 10 countries

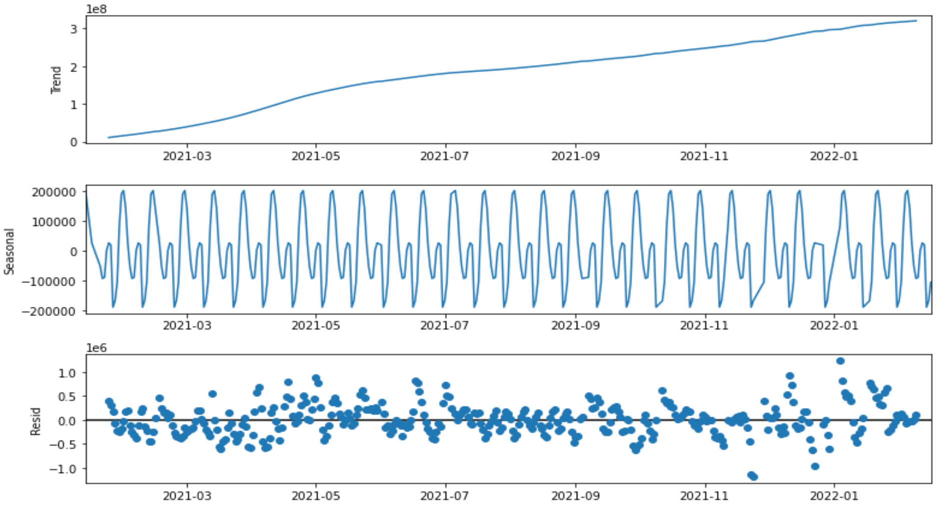

Figure 3. The ARIMA decomposition

We must make sure that our model's residuals are uncorrelated, normally distributed, and zero-mean. If not, it means that the model can be strengthened further, and we go through the same procedure using the residuals. Figure 4 demonstrates the normal distribution. Based on the following, our model diagnostics in this scenario indicate that the model residuals are normally distributed:
- The residuals on the Figure 4(b) KDE plot almost closely resemble the normal distribution.
- The ordered distribution of residuals (blue dots) as shown in the Figure 4(c) follows the linear trend of the samples selected from a standard normal distribution with N(0, 1). Once more, this strongly suggests that the residuals are normally distributed.
- The residuals over time (Figure 4(a)) do not appear to show any discernible seasonality and merely appear to be noise. This is supported by the autocorrelation (also known as a correlogram) Figure 4(d), which illustrates the time series residuals' low correlation with their lagged counterparts.

These findings, along with the absence of spikes outside the insignificant zone of correlogram plots, lead us to believe that the residuals are random and lack any information or juice, and our model generates a satisfactory fit that could aid in our understanding of the time series data and future value forecasting. Our model seems to be operating without any problems. Figure 5 shows the forecasting result for the next (90) days ahead for the vaccinations in the world.

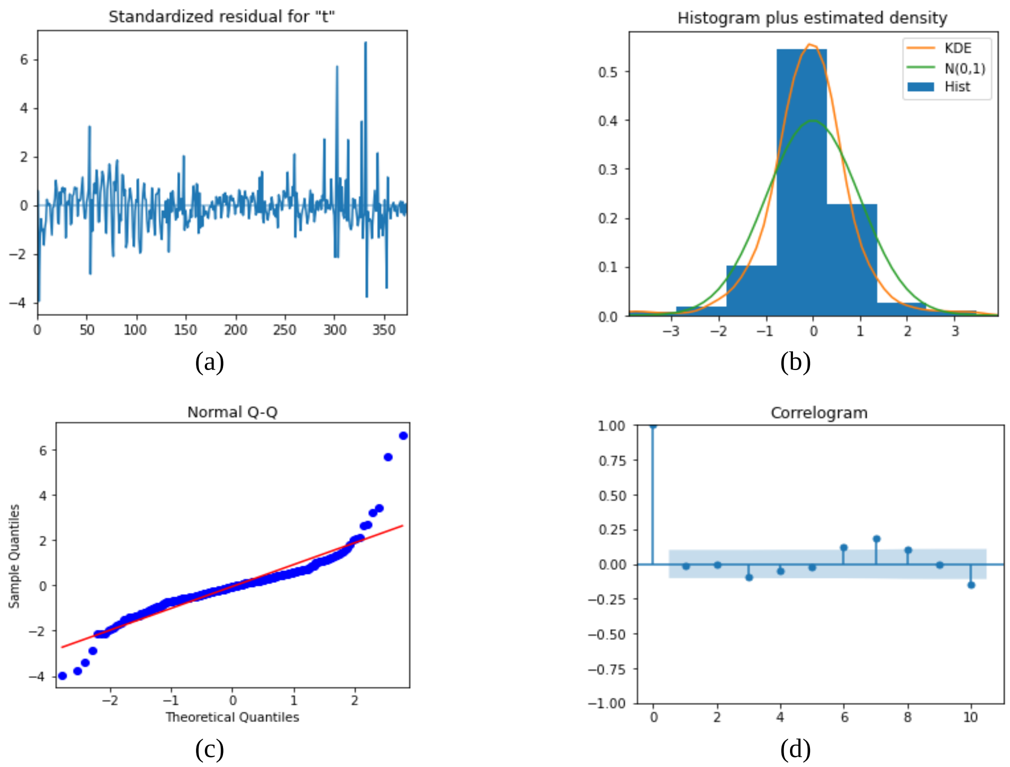

Figure 4. Residual plots (a) standardized residual for “t”, (b) histogram plus estimated density, (c) theoretical quantiles, and (d) autocorrelation

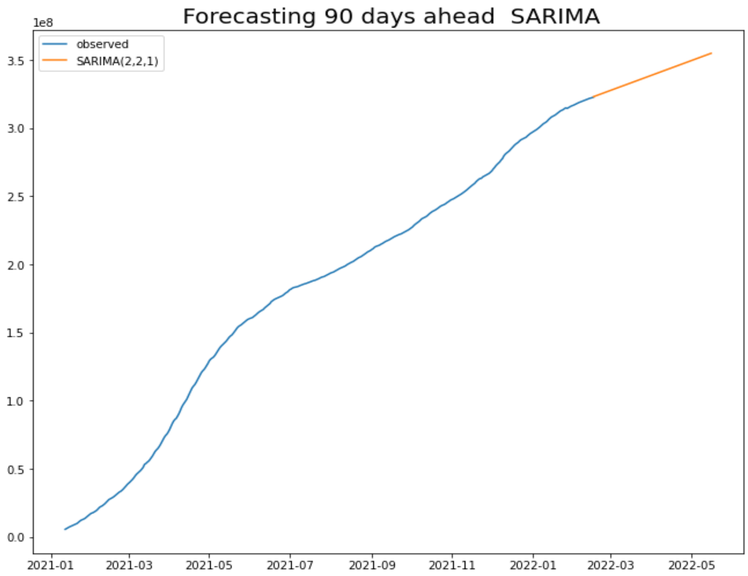

Figure 5. Forecasting 90 days ahead

3.2. Prediction for tweets

At this stage, the number of tweets that were written during the vaccination period and whose dates matched the date of the vaccinations will be processed. Figure 6 represents the number of tweets with dates corresponding to the dates of vaccinations. Best model: ARIMA (6,0,0) (0,0,0) [0], total fit time: 3.485 seconds. Automated model selection choses SARIMA (6,0,0) as best model based on AIC. Prob(Q)=0.00<0.05. the null hypothesis is rejected that the residuals are correlated so the residuals are correlated. Prob (JB)=0.00<0.05. the null hypothesis is rejected that the residuals are normally distributed. Therefore, the residuals are normally distributed.

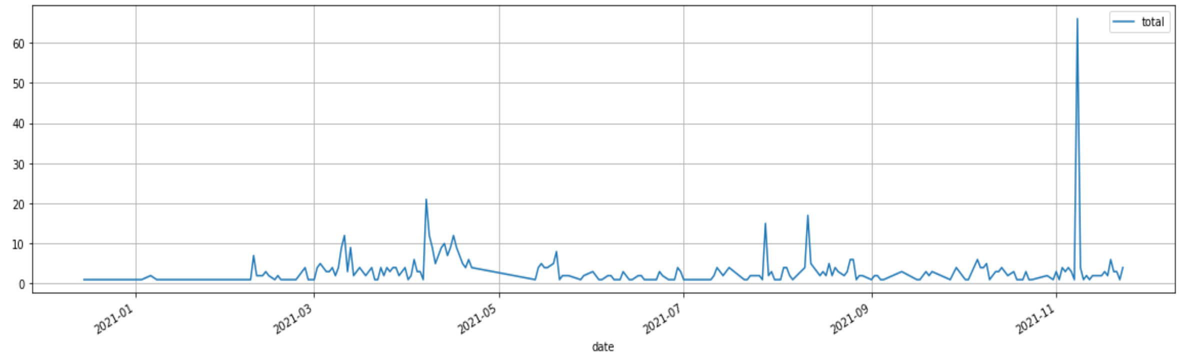

Figure 6. The tweets number

4. RESULTS AND DISCUSSION

The residuals of an ideal model should consist of uncorrelated white Gaussian noise centred at zero. By studying the above charts with this in mind, we can determine whether or not our model is accurate. For sample-based forecasting, we utilize the get prediction method using the last days of training data as validation data. Then, we utilize sklearn. Metrics' mean absolute error and mean absolute percentage error to calculate MAE and MAPE for the model. Table 1 displays the metric values used to assess the proposed model. Figure 5 shows the forecasting result for the next (90) days ahead for the vaccinations in the world.

The ideal model of tweets is chosen utilize sklearn. After studying the above charts with this in mind, the model can be determined if it accurate or not. For sample-based forecasting, the last days of training data is utilize as validation data for the prediction method. Then, the metrics mean absolute error and mean absolute percentage error are used to calculate MAE and MAPE for the model. Table 2 displays the metric values used to assess the proposed model. Figure 7 shows the prediction result for the next (90) days ahead for the negative tweets.

Table 1. The metrics value for the total number of vaccinations

Metric	SARIMA (2,2,1)
MAE	1.478688e+07
MAPE	4.900000e-02

Table 2. The metrics value for the number of tweets

Metrics	SARIMA (2,2,1)
MAE	1.0221
MAPE	0.919

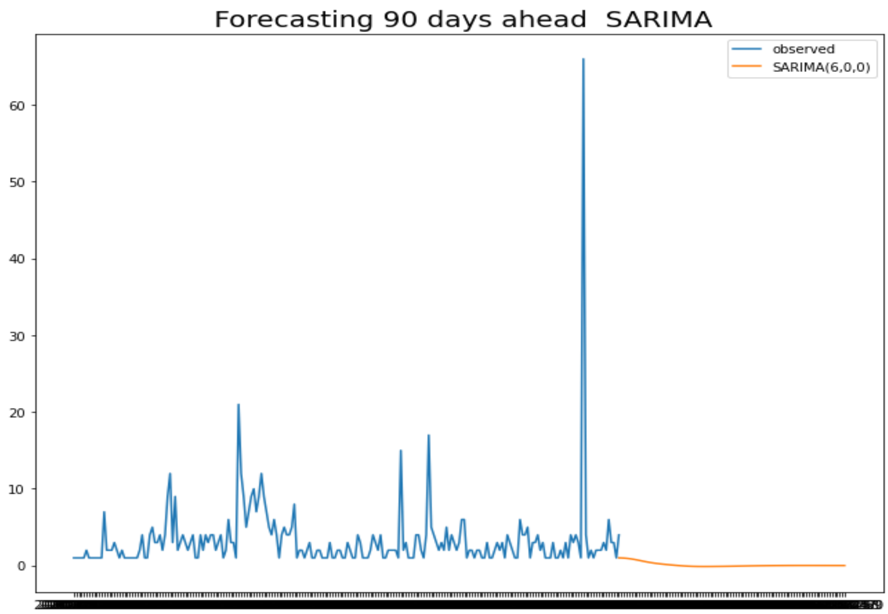

Figure 7. The prediction of 90 day ahead for the negative tweets

Post analysis: after the results of predicting the numbers of vaccinated people appeared and it appeared that the numbers were constantly increasing. This shows that there is more demand for taking the vaccine, which means that people's confidence in the vaccine began to increase, and what proves this is the emergence of the results of predicting the number of negative tweets, which was constantly decreasing with time. It means that there is an inverse relationship between the number of vaccinated people and the number of negative tweets about the vaccine.

5. CONCLUSION

Introduces ARIMA models and their variants: SARIMA and ARIMAX, which employ external data (exogenous inputs) to enhance the performance of the ARIMA model. The Box-Jenkins approach was used to identify the optimal model for a portion of the data set (time series of vaccinations). Important time series features, such as stationarity and seasonality, are recognized as the first step. Once the model has identified an acceptable solution, it is used to forecast in a sample, i.e., it is applied to a subset of the training data as validation data. After then, the projection extends 90 days beyond the sample period. The findings demonstrated a high degree of accuracy (a very low error rate) for the proposed model, which is optimistic for the future use of this model for forecasting the preparation.

REFERENCES

[1] A. S. Imran, S. M. Daudpota, Z. Kastrati, and R. Batra, "Cross-Cultural Polarity and Emotion Detection Using Sentiment Analysis and Deep Learning on COVID-19 Related Tweets," *IEEE Access*, vol. 8, pp. 181074–181090, 2020, doi: 10.1109/ACCESS.2020.3027350.

[2] K. H. Alhussayni and E. S. Alshamery, "Automated COVID-19 Dialogue System Using a New Deep Learning Network," *Periodicals of Engineering and Natural Sciences (PEN)*, vol. 9, no. 2, Apr. 2021, doi: 10.21533/pen.v9i2.1862.

[3] P. Cihan, "Forecasting fully vaccinated people against COVID-19 and examining future vaccination rate for herd immunity in the US, Asia, Europe, Africa, South America, and the World," *Appl Soft Comput*, vol. 111, p. 107708, Nov. 2021, doi: 10.1016/j.asoc.2021.107708.

[4] K. N. Alam *et al.*, "Deep Learning-Based Sentiment Analysis of COVID-19 Vaccination Responses from Twitter Data," *Computational and Mathematical Methods in Medicine*, vol. 2021, p. e4321131, Dec. 2021, doi: 10.1155/2021/4321131.

[5] N. S. Sattar and S. Arifuzzaman, "COVID-19 Vaccination Awareness and Aftermath: Public Sentiment Analysis on Twitter Data and Vaccinated Population Prediction in the USA," *Applied Sciences*, vol. 11, no. 13, Jan. 2021, doi: 10.3390/app11136128.

[6] M. Mahdikhani, "Predicting the popularity of tweets by analyzing public opinion and emotions in different stages of Covid-19 pandemic," *International Journal of Information Management Data Insights*, vol. 2, no. 1, p. 100053, Apr. 2022, doi: 10.1016/j.jjimei.2021.100053.

[7] R. Bose, S. Aithal, and S. Roy, "Sentiment Analysis on the Basis of Tweeter Comments of Application of Drugs by Customary Language Toolkit and TextBlob Opinions of Distinct Countries," *International Journal of Emerging Trends in Engineering Research*, vol. 8, pp. 3684–3696, Jul. 2020, doi: 10.30534/ijeter/2020/129872020.

[8] H. K. Obayed, F. S. Al-Turaihi, and K. H. Alhussayni, "Sentiment classification of user's reviews on drugs based on global vectors for word representation and bidirectional long short-term memory recurrent neural network," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 23, no. 1, Jul. 2021, doi: 10.11591/ijeecs.v23.i1.pp345-353.

[9] H. Jelodar *et al.*, "A NLP framework based on meaningful latent-topic detection and sentiment analysis via fuzzy lattice reasoning on youtube comments," *Multimedia Tools Appl.*, vol. 80, no. 3, pp. 4155–4181, Jan. 2021, doi: 10.1007/s11042-020-09755-z.

[10] M. Kim, "Prediction of COVID-19 Confirmed Cases after Vaccination: Based on Statistical and Deep Learning Models," *SciMedicine Journal*, vol. 3, no. 2, Jun. 2021, doi: 10.28991/SciMedJ-2021-0302-7.

[11] S. F. Ardabili *et al.*, "COVID-19 Outbreak Prediction with Machine Learning," *Algorithms*, vol. 13, no. 10, Oct. 2020, doi: 10.3390/a13100249.

[12] L. J. Muhammad, E. A. Algehyne, S. S. Usman, A. Ahmad, C. Chakraborty, and I. A. Mohammed, "Supervised Machine Learning Models for Prediction of COVID-19 Infection using Epidemiology Dataset," *SN COMPUT. SCI.*, vol. 2, no. 1, p. 11, Nov. 2020, doi: 10.1007/s42979-020-00394-7.

[13] L. A. Amar, A. A. Taha, and M. Y. Mohamed, "Prediction of the final size for COVID-19 epidemic using machine learning: A case study of Egypt," *Infectious Disease Modelling*, vol. 5, pp. 622–634, Jan. 2020, doi: 10.1016/j.idm.2020.08.008.

[14] P. G. J and P. K. H. R, "Sentiment Analysis:Textblob For Decision Making," *International Journal of Scientific Research & Engineering Trends*, vol. 7, no. 2, pp. 1097–1099, 2021.

[15] D. Hazarika, G. Konwar, S. Deb, and D. J. Bora, "Sentiment Analysis on Twitter by Using TextBlob for Natural Language Processing," in *Annals of Computer Science and Information Systems*, 2020, vol. 24, pp. 63–67, doi: 10.15439/2020KM20.

[16] S. Sohangir, N. Petty, and D. Wang, "Financial Sentiment Lexicon Analysis," in *2018 IEEE 12th International Conference on Semantic Computing (ICSC)*, Jan. 2018, pp. 286–289, doi: 10.1109/ICSC.2018.00052.

[17] V. Bonta, N. Kumaresh, and N. Janardhan, "A Comprehensive Study on Lexicon Based Approaches for Sentiment Analysis," *Asian Journal of Computer Science and Technology*, vol. 8, no. S2, pp. 1–6, Mar. 2019, doi: 10.51983/ajcst-2019.8.S2.2037.

[18] H. Khalid, N. Al-A'araji, and E. Al-Shamery, "Examination and Forecasting of Drug consumption Based on Recurrent Deep Learning," *International Journal of Recent Technology and Engineering (IJRTE)*, vol. 8, no. 2S10, pp. 414–420, Oct. 2019, doi: 10.35940/ijrte.B1069.0982S1019.

[19] S. Chae, S. Kwon, and D. Lee, "Predicting Infectious Disease Using Deep Learning and Big Data," *International Journal of Environmental Research and Public Health*, vol. 15, no. 8, Aug. 2018, doi: 10.3390/ijerph15081596.

[20] L. Gaspareniene and R. Remeikiene, "Arima model for predicting the development of the price of gold: european approach," *Ekonomicko-manazerske spektrum*, vol. 14, no. 1, pp. 87–96, Jun. 2020, doi: 10.26552/ems.2020.1.87-96.

[21] A. Khakharia *et al.*, "Outbreak Prediction of COVID-19 for Dense and Populated Countries Using Machine Learning," *Annals of Data Science*, vol. 8, no. 1, pp. 1–19, Mar. 2021, doi: 10.1007/s40745-020-00314-9.

[22] S. I. Alzahrani, I. A. Aljamaan, and E. A. Al-Fakih, "Forecasting the spread of the COVID-19 pandemic in Saudi Arabia using ARIMA prediction model under current public health interventions," *Journal of Infection and Public Health*, vol. 13, no. 7, pp. 914–919, Jul. 2020, doi: 10.1016/j.jiph.2020.06.001.

[23] R. J. Hyndman, "ARIMA processes-Rob J Hyndman," *yumpu.com*, 2001. https://www.yumpu.com/en/document/view/29441084/arima-processes-rob-j-hyndman.

[24] H. K. Obayes, "Suggested Approach to Embedded Playfair Cipher Message in Digital Image," *Int. Journal of Engineering Research and Application*, vol. 3, no. 5, pp. 710–714, Oct. 2013.

[25] A. de Myttenaere, B. Golden, B. L. Grand, and F. Rossi, "Mean Absolute Percentage Error for regression models," *Neurocomputing*, vol. 192, pp. 38–48, Jun. 2016, doi: 10.1016/j.neucom.2015.12.114.

BIOGRAPHIES OF AUTHORS

Hadab Khalid Obayes she is assistant professor at University of Babylon, Babel, Iraq. She received her Ph.D degree from University of Babylon, Iraq, in 2020. She published a number of papers in international journals and referred conferences. Her current research areas are AI, machine learning, neural networks, deep learning, natural language processing, and image processing. She can be conctacted at email: Hedhab@uobabylon.edu.iq.

Khaldoon Hasan Alhussayni received the B.Sc. degrees in Computer Science from the University of Babylon, Iraq, in 2008. After completing his BSc, he worked as a programmer at the Department of Computer Center, the University of Babylon in 2009. He received the MSc degrees in Computer Science and Computer Engineering from the Tula State University, Russian, in 2014. Secince Nov 2016. He entered the Tomsk state university (TSU), Department of computer science theory as a Ph.D. student in specialty Mathematical and software of computers, complexes and computer networks. His main research interests are machine learning, neural networks, deep learning, dialogue systems and natural language processing. He can be conctacted at email: alhussayni83@gmail.com.

Saba Mohammed Hussain she completed B.Sc. in Computer Sciences in 1998 from the college of sciences at University of Babylon, Babil, Iraq. Received his M.Sc. (master) in computer sciences in the year 2011 from the college of computer sciences at the Babylon University (BU), Babylon, Iraq. She received her PhD degree from University of Babylon, Iraq, in 2020. She has worked as a lecturer in Information Technology in Babylon Universities in the areas of artificial intelligent, information system and data base. She currently works as a lecturer in the University of Babylon in information Technology. Her research interests machine learning, social network algorithms, and datamining. She can be conctacted at email: saba.alshebeeb@uobabylon.edu.iq.