DOI: 10.24996/ijs.2024.65.1.37

ISSN: 0067-2904

Predicting COVID-19 in Iraq using Frequent Weighting for Polynomial Regression in Optimization Curve Fitting

Mohammed AL-Mukhtar *, Ammar S. Al-Zubaidi, Mustafa N. Albadri
Computer Center, University of Baghdad, Baghdad, Iraq

Received: 27/8/2022 Accepted: 18/2/2023 Published: 30/1/2024

Abstract

The worldwide pandemic Coronavirus (Covid-19) is a new viral disease that spreads mostly through nasal discharge and saliva from the lips while coughing or sneezing. This highly infectious disease spreads quickly and can overwhelm healthcare systems if not controlled. However, the employment of machine learning algorithms to monitor analytical data has a substantial influence on the speed of decision-making in some government entities. ML algorithms trained on labeled patients' symptoms cannot discriminate between diverse types of diseases such as COVID-19. Cough, fever, headache, sore throat, and shortness of breath were common symptoms of many bacterial and viral diseases.

This research focused on the numerous tendencies and projected expansion of the Iraq pandemic to encourage people and governments to take preventive measures. This work is an established basic benchmark for demonstrating machine learning's capabilities for pandemic prediction.

The suggested approach for forecasting the number of COVID-19 cases can assist governments in taking safeguards to avoid the disease's spread. We have demonstrated the effectiveness of our strategy using publicly available datasets and models. A polynomial network is trained on this premise, and the parameters are optimized using frequent weighting. When compared to linear models, the polynomial model predicts better and is more effective in forecasting COVID-19 new confirmed cases. As well, it aims to analyze the spread of COVID-19 in Iraq and optimize polynomial regression. In time series-based models, curve fitting using frequent weighting to implement models such as linear regression and polynomial regression is utilized to estimate the new daily infection number. The datasets were collected from March 13, 2020, to December 12, 2021. The continuous COVID-19 pandemic puts both human lives and the economy at risk. If AI could forecast the next daily hospitalization number, it may be a useful tool in combating this pandemic sickness.

Keywords: COVID-19, Machine learning, linear regression, polynomial regression.

توقع COVID-19 في العراق باستعمال ترجيح متكرر للانحدار متعدد الحدود في ملاءمة منحنى التحسين

محمد المختار*, عمار الزبيدي, مصطفى البدري
مركز الحاسبة, جامعة بغداد, بغداد, العراق

*Email: mohammed.abdul@cc.uobaghdad.edu.iq

الخلاصة

فيروس كورونا الوبائي العالمي (كوفيد -19) هو مرض فيروسي جديد ينتشر في الغالب عن طريق إفرازات الأنف واللعاب من الشفتين أثناء السعال أو العطس. ينتشر هذا المرض شديد العدوى بسرعة ويمكن أن يطغى على أنظمة الرعاية الصحية إذا لم يتم السيطرة عليه. ومع ذلك ، فإن توظيف خوارزميات التعلم الآلي لمراقبة البيانات التحليلية له تأثير كبير على سرعة اتخاذ القرار في بعض الجهات الحكومية. لا تستطيع خوارزميات التعليم الالي المدربة على أعراض المرضى المصنفين التمييز بين أنواع مختلفة من الأمراض مثل COVID-19. لأن هناك العديد من الأمراض البكتيرية والفيروسية لها نفس الأعراض مثل السعال والحمى والصداع والتهاب الحلق وضيق التنفس. ركز هذا البحث على الاتجاهات العديدة والتوسع المتوقع لوباء العراق للناس والحكومات لاتخاذ تدابير وقائية. يعد هذا العمل معيارًا أساسيًا ثابتًا لإظهار قدرات التعلم الآلي للتنبؤ بالأوبئة. يمكن للنهج المقترح للتنبؤ بعدد حالات COVID-19 أن يساعد الحكومات في اتخاذ إجراءات وقائية لتجنب انتشار المرض. لقد أظهرنا فعالية استراتيجيتنا باستخدام مجموعات البيانات والنماذج المتاحة للجمهور. يتم تدريب شبكة متعددة الحدود على هذه الفرضية ، ويتم تحسين المعلمات باستخدام ترجيح متكرر. مقارنة بالنماذج الخطية ، تحقق كثير الحدود نتائج أفضل للتنبؤ وأكثر فاعلية في التنبؤ بحالات COVID-19 المؤكدة الجديدة. بالإضافة إلى ذلك ، يهدف إلى تحليل انتشار COVID 19 في العراق وتحسين ملاءمة منحنى الانحدار متعدد الحدود باستخدام ترجيح متكرر لتنفيذ النماذج مثل الانحدار الخطي والانحدار متعدد الحدود. يتم استخدام هذه النماذج في النماذج المستندة إلى السلاسل الزمنية لتقدير عدد الإصابة اليومي الجديد. تم جمع مجموعات البيانات من 13 مارس 2020 إلى 12 ديسمبر 2021. وباء COVID-19 المستمر يعرض حياة البشر والاقتصاد للخطر. إذا كان بإمكان الذكاء الاصطناعي التنبؤ برقم الاستشفاء اليومي التالي ، فقد يكون أداة مفيدة في مكافحة هذا المرض الوبائي.

1. Introduction

COVID-19), originated in the city of Wuhan, China, in late 2019. Then it spread so fast all over China in the next few months. On March 11, 2020, the World Health Organization declared the term COVID-19 a pandemic and started a wide campaign to warn the world about it [1]. COVID-19 is one of the diseases of zoonotic origin; it is transmitted from human to human by different routes that include the transmission of small droplets from the nose or mouth of infected persons through sneezing, coughing, or speaking. [2].

COVID‑19 spreads between people more easily than influenza. People are very infectious when they show symptoms, but they might be infectious for up to a couple of days before symptoms appear. They are still infectious for an estimated 7 to 12 days in mild cases and for an average of a couple of weeks in tough cases. An individual can also transmit COVID-19 without showing any symptoms (asymptomatic transmission) [3].

This virus has caused widespread suffering all over the world and substantial disruptions in global economies. There is no approved immunization or cure for fighting it at that time. When humans are nearby, COVID-19 engendering proceeds quickly [4].

COVID-19 has had a significant impact on our lives. Social relationships, education, work, and the economy have come to a standstill. People today only focus on what is necessary to survive. Governments and other organizations have had to take extraordinary steps to fight the disease. To stop the disease's spread, practically all the countries have imposed either partial or full lockdowns in the affected regions [5].
Since there is no concrete solution until that time for dispatching this contamination, the authorities of all countries are working hard to achieve careful steps that can minimize the

spread. Procedures like face masks, sanitizers, and social distancing are approved to break the virus's spread [6].

The COVID-19 problem can be better understood and examined by machine learning, which recognizes data patterns and uses them to make predictions or judgments on its own.
Machine learning can be considered a resource in the medical services sector with the capacity to process enormous datasets beyond the capabilities of human personalities, and the knowledge generated from it aids clinicians in planning and considering getting appropriate treatment [7]. Machine learning is a subset of artificial intelligence that enables computers to behave and make decisions in the same way that humans do. It also allows us to make more accurate predictions and boosts our efficiency.

In this model, the authors are checking the development of COVID-19. Investigation and predictions are done using linear and polynomial regression. The suggested expectation model makes sure that it accurately predicts the outcome of this outbreak situation so that severe financial loss, network spread, and the degree of social segregation among individuals can all be used as indicators and pre-cruise decisions can be made as well [8].

This approach will guarantee that the administration produces preventive assessments dependent on our subsequent research for predicting the presence of this virus in the future.
The study focuses on COVID-19 disease detection. In the first step, sample data was pre-processed based on specified characteristics, followed by a classification approach using some of the following machine learning algorithms in Python.

2. literature review

Since COVID-19's emergence and global growth, it has overtaken both developed and developing nations. A lot of work has gone into disseminating exploration papers on the various COVID-19 components. Therefore, studies of vaccination, medication therapy regimens, and the prediction of future infected deaths and recovered cases utilizing a variety of forecasting approaches are done in many research papers.

The authors of a recent article in [9] conducted an online survey to gather information in Jordan. This data was used as an input for several statistical (Logistic Regression, LR) and machine learning-based prediction models, including Support Vector Machine (SVM) and Multi-Layer Perceptron (MPL). These algorithms were utilized to predict prospective COVID-19 patients based on their symptoms and indicators. In comparison to other models, the MLP has demonstrated the highest accuracy (91.62%). The SVM, on the other hand, displayed the highest precision (91.67%).

Machine learning (ML) is being used as a persuasive method to show flare-ups [4]. The COVID-19 data set from several online resources has been assessed using the polynomial, linear, ridge, polynomial ridge, and SVM regression models. Comprehensive tests were run, and the metrics medium absolute error (MAE), R2 score, mean square error (MSE), and root mean square error (RMSE) were used to evaluate each test. Early on, accurate validation and data analysis can help with healing and illness prevention.

The authors created supervised ML models for the COVID-19 virus in this work [10]. They utilized an epidemiologically labeled dataset for negative and positive COVID-19 cases in Mexico. The prediction techniques employed are SVM, decision trees, logistic regression, and naive Bayes. Before creating the models, the correlation coefficient analysis of the different independent and dependent features was performed to ascertain a strong association between each independent and dependent feature of the dataset. The models were trained on 80% of the training dataset, and they were tested on the residual 20%. The decision tree has the maximum accuracy of 94.99%, according to the results of the models' performance evaluation, while the SVM has the lowest accuracy.

In this study [11], the authors used deep learning models with an autoencoder-based approach along with machine learning with hyperparameter tuning to estimate the impact of various features on the spread of the disease and forecast the chances of survival for the patients who were placed in quarantine. The study's calibrated model is based on positive coronavirus infection cases and includes an analysis of several factors that are useful in examining temporal patterns in the current situation and fatality cases brought on by coronavirus. The analysis highlights pivotal moments in the progression of the outbreak, demonstrating the viability of models powered by artificial intelligence and deep learning in presenting a quantitative picture of the pandemic breakout.

This study [12] offers a meta-analysis of the global pandemic COVID-19 forecast trend based on artificial intelligence (AI). SVM, Naive Bayes, and Linear Regression. Here, potent ML techniques were used in a real-time series dataset that contained information on the confirmed deaths, recoveries, active cases, and confirmed cases of the COVID-19 pandemic worldwide. Additionally, statistical analysis has been done to depict different features of the Top 20 Coronavirus-affected Countries and the COVID-19 Observed Symptoms. Naive Bayes, one of three machine learning methods used, showed promise in predicting COVID-19 future trends with lower MAE and MSE. The Naive Bayes technique's efficiency is demonstrated by the lower values of MAE and MSE. This study identified numerous trends and the potential expansion of the worldwide pandemic, highlighting the need for citizens and governments around the world to take proactive measures.

By examining the COVID-19 data, this paper [13] aimed to predict the COVID-19 active, death, and cure rates in India. For predicting the death rate, active rate, and cure rate, three ML models are available: reinforced SVM, linear regression, and the Prophet Forecasting Model. In comparison to linear regression and SVM, the Prophet Forecasting Model provides the most accurate predictions of the death rate, active rate, and cure rate for large, uncertain, and sparse data sets.

A reliable worldwide data set is gathered from the WHO's daily statistics [14]. This research states a correlation between the total number of confirmed, active, deceased, and positive cases. On the data set, multiple linear regression and linear regression techniques are used to forecast the trend of the affected cases. The scores of the model R2 tend to be 0.99 and 1.0, correspondingly, which shows a strong prediction model to anticipate the active cases for the following day. An evaluation of the linear regression and linear regression models is conducted. As of July 2020, the multiple linear regression model predicts that there will be 52,290 active cases in India by the 15th of August and 9,358 active cases in Odisha. if things keep going the way they are. A high correlation factor (positive, deceased, or recovered) determines the relationship between the independent and dependent (active) variables.

Using the common Susceptible-Infected-Removed pandemic model, this study seeks to estimate the fundamental reproductive number for COVID-19 in the Kurdistan Region of Iraq. The 4^{th}-order Runge-Kutta method was used to formulate and solve a system of nonlinear differential equations numerically. By using the least squares method to match the curves between the real daily data and the numerical solution, the reproductive numbers R0 were determined. For the analysis, data were collected for 165 days, from 1 March to 12 August 2020, on a population of 5.2 million people. It was determined that the R0 value fluctuated throughout the outbreak, averaging 1.33, and that the greatest number of infection cases, or roughly 540,000, will occur on the fifth of November 2020. The disease will then stop spreading once there are only roughly 3.2 million susceptible individuals left. The total number of people who have been eliminated will be close to 1.5 million [15].

3. Methodology

The study was conducted using two machine learning algorithms: linear regression and optimized polynomial regression. The approach used for this analysis is outlined in the parts that follow.

3.1 dataset collection.

Our data set was collected from the official Iraqi Ministry of Health repository. This repository provides daily updates on confirmed, death, and recovered cases from across the country, along with detailed statistics [16].

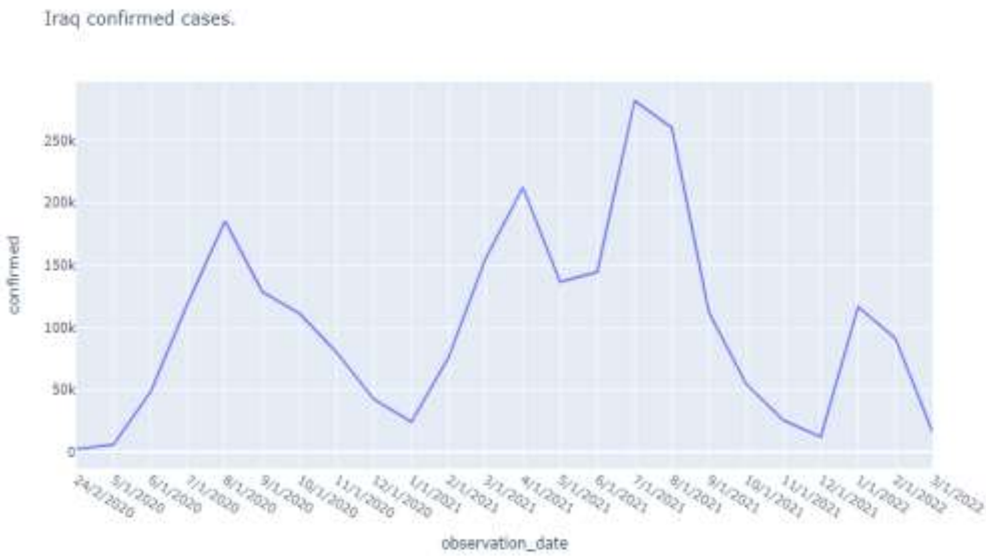


Figure 1: (a) Iraq confirmed cases plot

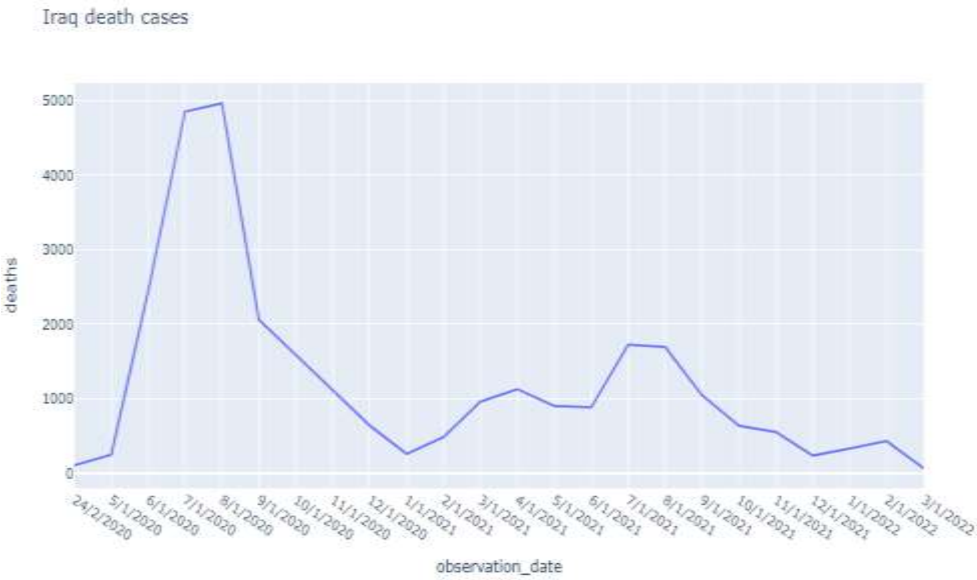


Figure 1**:** (b) Iraq death cases plot

Figure 1(A) shows the number of confirmed cases; in March 2020, the first cases started to appear in Iraq. The government started a total lockdown at the end of March to break down the spread of COVID; however, in June 2020, the cases touched 1000 cases a day and things started to be a tragedy. In September 2020, we will reach the peak of the first wave, with over 4000 confirmed cases per day. Also, the death rate was relatively high compared to Iraq's neighbors in the area, like KSA, Jordan, and Kuwait. Poor health infrastructure and a large number of people who did not comply with the lockdown due to not using precautionary measures such as masks and sanitizing increased ER and death cases. The second wave started in March 2021, and new cases reached 8,000 in April 2021 and 13,000 in August 2021. However, the death rate dropped because of the vaccine. Many people got their first dose in April 2021 and their second dose by July of the same year. COVID protocol procedures became more mature and efficient, which helped decrease ER cases and death rates. At the beginning of 2022, when the 3rd wave starts rising in Iraq, the new confirmed cases reached 8K for a short period, aside from the very low ER and death cases because of the huge number of vaccinated people. New cases started to diminish in March 2022, but our model shows that there was a rise in new cases in June.

Figure (1(B)) The death rate graph shows that the first wave has the most death cases, despite the fact that those new cases weren't too high compared to the second and third waves. This was because we don't have much information about the new pandemic curing protocols. As well as the Iraqi health infrastructure being so bad, there was a shortage of beds, equipment, and well-trained personnel. The second wave had lower death rates despite the rise in new cases. That is because Iraq learned from the first wave, and the curing protocols became more mature. Also, the new vaccine made ER visitors less frequent than before, which made things easier for healthcare workers because hospitals became less crowded, which made them take care of more in-patients. Because many people had received two vaccine doses in addition to the third, the death rate in the third wave was lower than in the previous two waves.

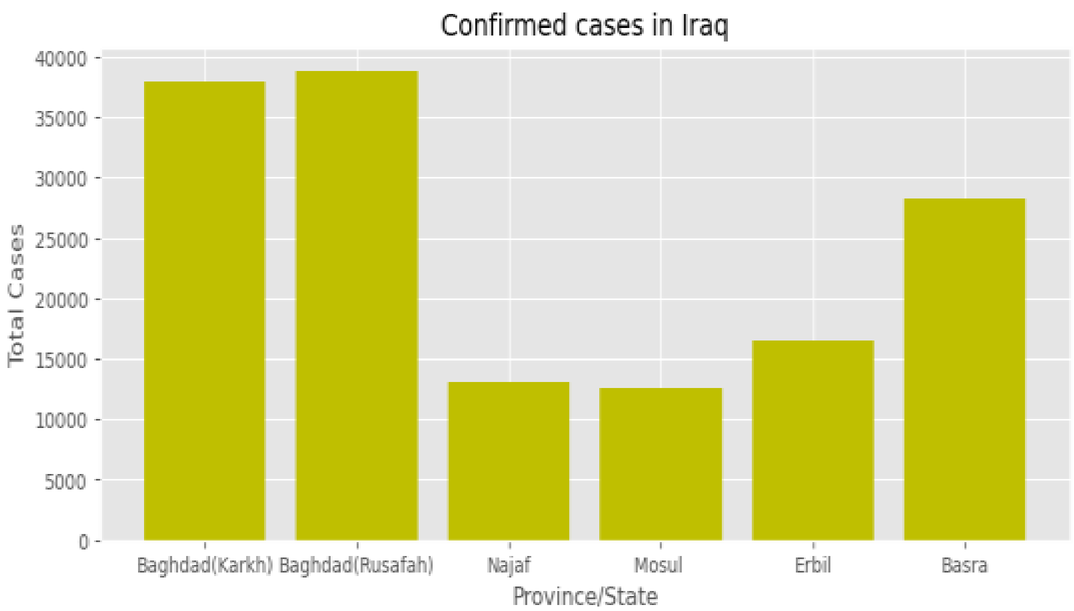


Figure 2: comparative infection of confirmed cases in some Iraqi provinces

The graph above shows that there is a difference between confirmed cases in many Iraqi provinces. Baghdad (Rusafah) had the most cases, followed by Baghdad (Kharkh). This can be explained because it had the most population (around 20% of the Iraqi population for both). Despite having the third largest population in Iraq, Mosul had the fewest cases of the six samples in the graph. This can be explained by the fact that many people there live in suburban areas where testing facilities are not nearby, so many of them didn't test themselves despite being positive for COVID-19.

3.2 Polynomial regression:

The polynomial technique approximates a nonlinear relationship between the dependent value of x and the conditional mean of the independent variable y [17].

$$Y_i = b_0 + b_1 x_1 + b_2 x_1^2 + b_3 x_1^2 ... + b_n x_1^n \quad (1)$$

Y is a dependent variable, and DV, $X_{1.}$ X_n is an Independent Variable (IV), b_0 intercept, $b_{1,}$ b_2 Coefficients, n number of observations.

As well as we can determine some Advantages and Disadvantages of Polynomial Regression:

A- Polynomial regression provides the best approximation of how the dependent and independent variables are related. It can accommodate a wide range of functions and suits a broad variety of curvatures.

B- The disadvantages of the results of the nonlinear analysis may be significantly impacted by one or two outliers in the data. These are excessively outlier-sensitive.

Generally, the main idea of polynomial distribution can be explained precisely as follows:

1- Binomial Distribution: The likelihood of a SUCCESS or FAILURE outcome in a repeated experiment or survey is known as the binomial distribution. Two potential outcomes are present in the binomial distribution (the prefix "bi" implies two, or twice). A coin flip, for instance, only has two outcomes: heads or tails, whereas taking a test has two outcomes: pass or fail.

2- Multinomial/Polynomial Distribution: Multi or Poly signifies a lot of things. The binomial distribution is a generalization of the multinomial distribution in probability theory. It may, for example, simulate the likelihood of each side's count after rolling a k-sided die n times. The multinomial distribution offers the chance of any specific combination of the total number of successes across all categories for n separate trials, each of which results in success for exactly one of k categories, with each category having a specified fixed success probability. The frequent weighting for the polynomial model was described precisely in Algorithm [1].

Algorithm [1] Optimization Curve fitting using frequent weighting for polynomial

Requirement: X_i *: input sequence of days from first case,* Y_i *: Number of cases for each day in* X_i

ε *: threshold parameter,* b_0: Intercept, $b_1 .. b_n$: Coefficients, n: number of observations

procedure *optimize curve fitting*

$w^0 \leftarrow$ Unit vector [1] × size (x)

for *iteration n from 0, step 1* ***do***

$f \leftarrow LM(input = X_i, target = Y_i, weights W^n)$

$d_i \leftarrow |f(x_i) - y_i| \forall i$

$$w_i^{n+1} \leftarrow \frac{\exp\left(1 - \frac{d_i^n - \tanh(d_i^n)}{\max_i d_i^n - \tanh(d_i^n)}\right)}{\sum_i \exp\left(1 - \frac{d_i^n - \tanh(d_i^n)}{\max_i d_i^n - \tanh(d_i^n)}\right)}$$

if $\sum_i \left| w_i^n - w_i^{n+1} \right| \le \varepsilon$ **then**

$$Y_i = b_0 + b_1 x_1 + b_2 x_1^2 + b_3 x_1^2 ... + b_n x_1^n$$

break

end for

end procedure

3.3 Linear regression:

Linear egression estimates the relationship values between two variables x and y, where y represents one dependent variable (DVs). ($x_1 \ldots x_n$) represents many independent variables (IVs), while b_0 is intercepted, and b_1, b_2 , b_3 are coefficients and n represent the number of observations. That is used in Eq. (2) for general linear processes [18].

$$Y_i = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 ... + b_n x_n \quad (2)$$

The relationship can be determined by fitting a linear equation to the dataset. When X is used as an explanatory variable, b determines how to generate a slope prediction.

The linear regression model interprets the prediction as a straight line between two variables. It tries to draw the closest line to the data by determining the slope and intercept, which define the line and minimize regression errors. There is only one (x) and one (y) variable in simple linear regression.

4. Experimental Results

Figure [3] depicts the comparative prediction for polynomial regression between Iraq and Kuwait, while Figure [4] depicts the result for Iraq in the Nissan month of 2022, and the predicted number of cases on 1-5-2022 was around 5230, which was close to the mister of health report on that day.

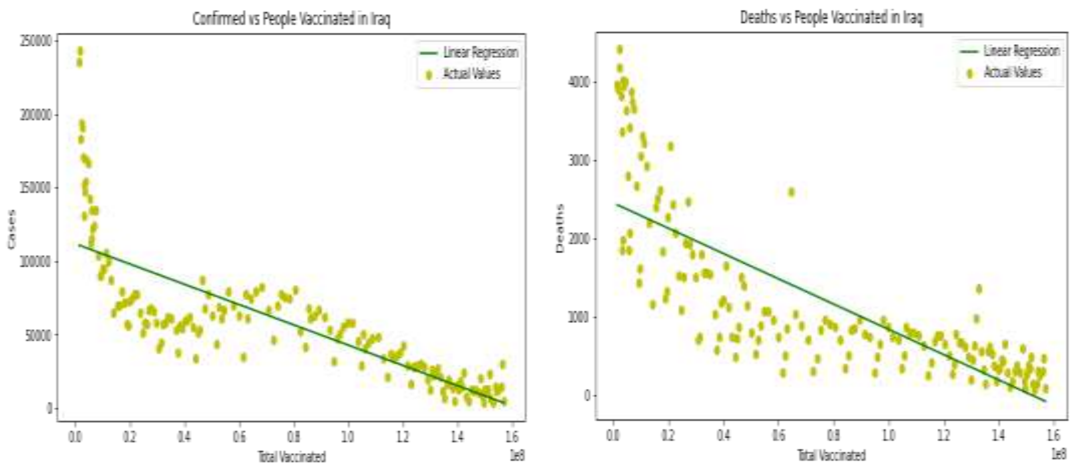


Figure 3: Predicted cases and confirm cases in Iraq related to total vaccinated people.

We perform linear regression on the data and display the best fit line. It is evident that the number of new cases each day and total vaccinations have a strong link, as shown in Figure 5. As the number of vaccines grows, the number of new cases each day tends to decrease. As a result, the number of confirmed COVID-19 cases in Iraq each day should continue to fall as the number of people vaccinated increases. The trend in the number of new fatalities each day vs.

total vaccinations is identical to that in the preceding graph. The line of best fit, as depicted, reveals a link between the mortality rate and total immunizations. The outliers do not appear to have had much of an impact on the linear regression, but they do appear to have had an impact on the model.

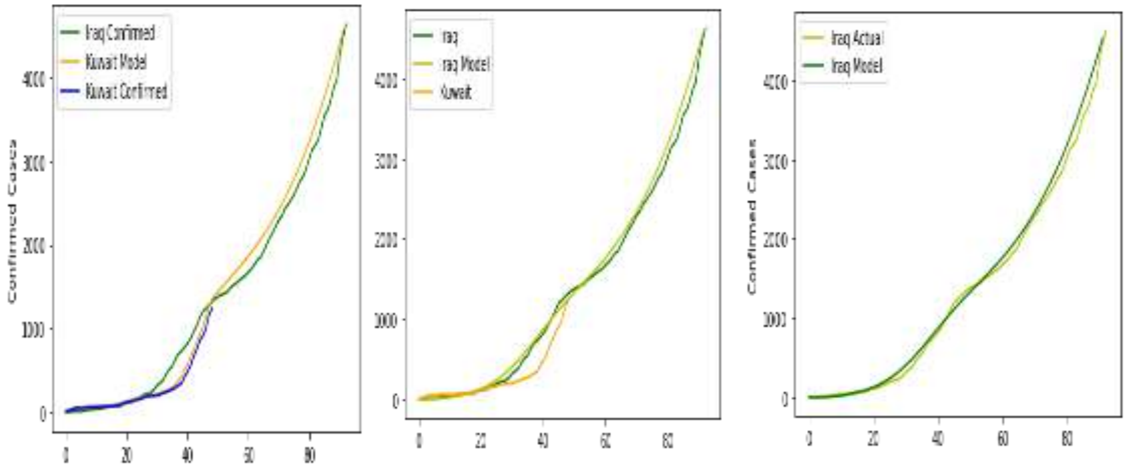


Figure 4: polynomial regression model comparative Iraq and Kuwait

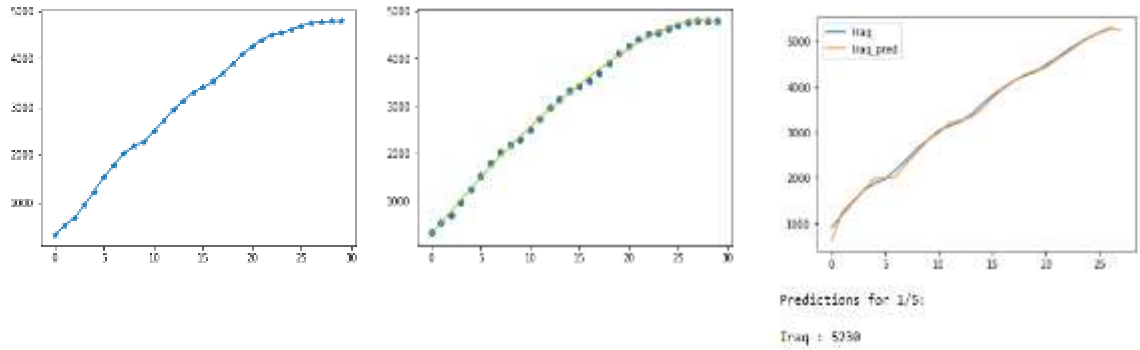


Figure 5: Polynomial Regression for April 2022

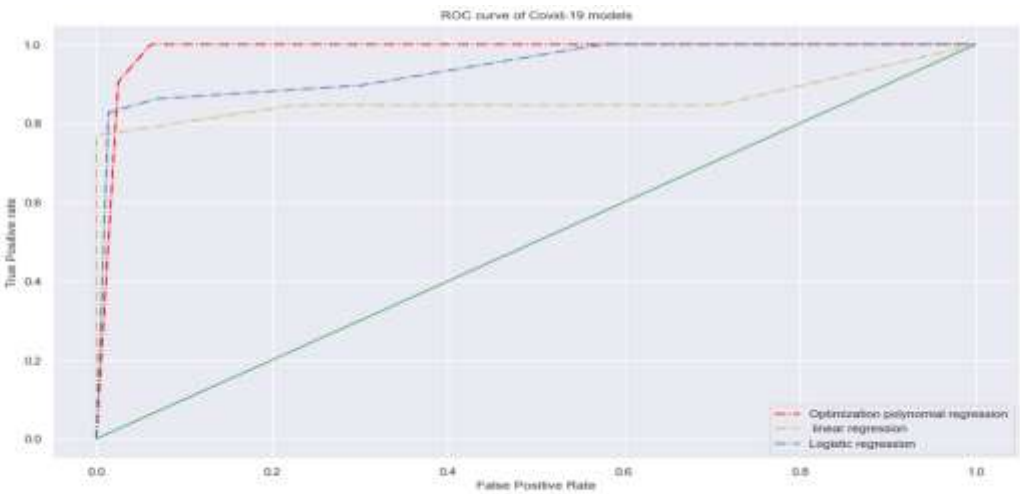


Figure 6: ROC curve of COVID 19 Models for Optimization polynomial regression

Evaluation of the model can be done through a ROC curve such as that shown in Figure 6, which compares the optimization PR with other standard ML techniques like linear regression and logistic regression.

However, the RMSE for four of our provinces: Baghdad, Erbil, Najaf, and Basra were reported in Table [1].

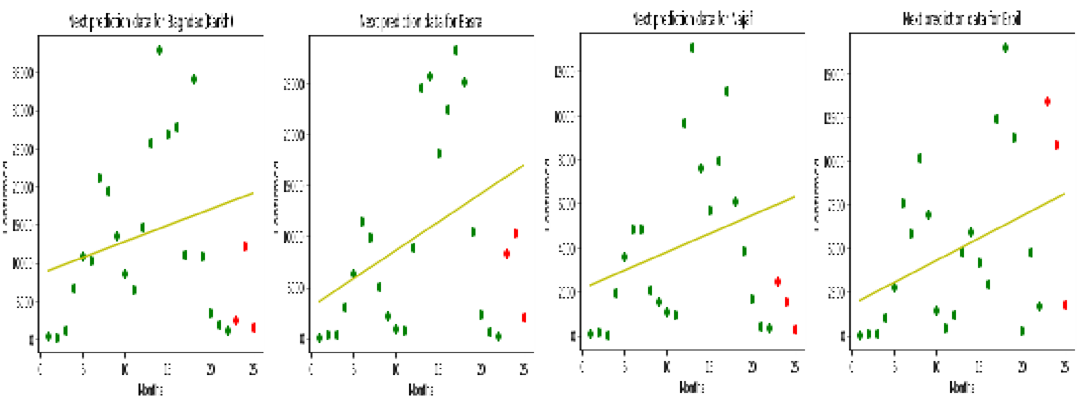
$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_i^2 - Y_i^2)} \tag{3}$$

Σ = summation ("add up"), (Y– Y) = differences, squared, N = sample size.

Eq. (3) The root mean square error (RMSE) is used to calculate the standard deviation of the residuals (prediction errors). The residuals show how far away the data points are from the regression line, and the RMSE shows how widely these residuals are spread out. In other words, it represents how tightly the data is clustered around the best fit line.

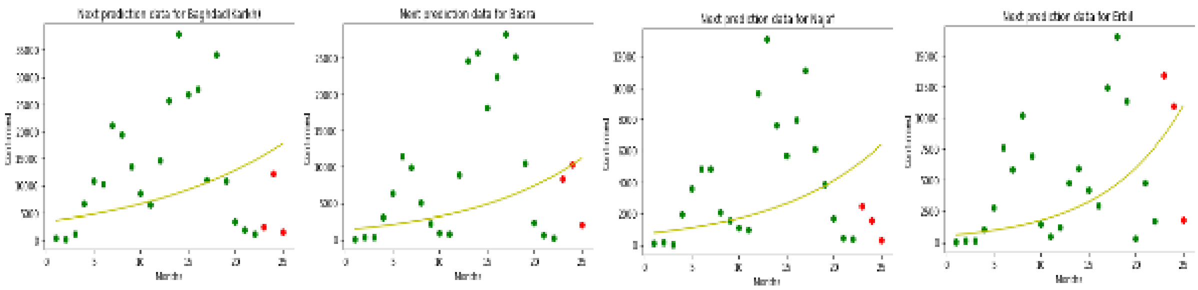
Table 1: RMSE comparative for our four provinces of confirmed cases of COVID-19

Provinces	Linear regression	logistic function	Polynomial Features	Polynomial Function	LassoLars Function	Scale the input data by a log function	Actual RMSE
Erbil	5263.93	6018.11	7244.84	7871.94	7005.55	1.157	7249.02
Baghdad	14189.09	12336.7	7189.33	11881.02	21971.08	2.44	7697.70
Najaf	4772.98	4571.12	1680.36	5941.27	7758.81	1.84	1468.30
Basra	10270.38	5327.58	7766.54	12750.94	5470.26	5.02	7697.70



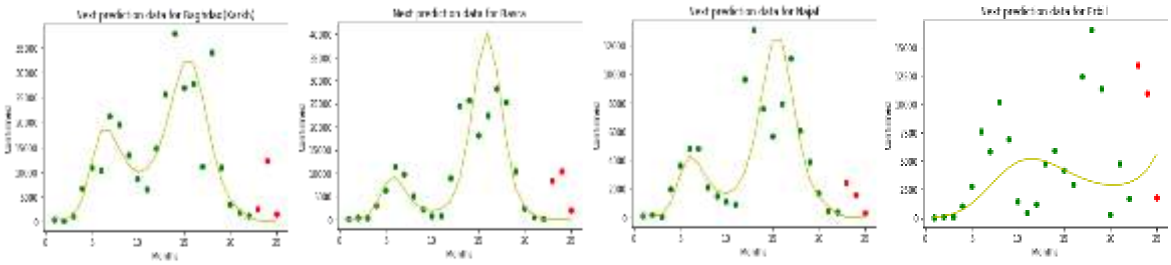
(a) **(b)** **(c)** **(d)**

Figure 7: fit using Linear Features function (a) Baghdad (Karkh) (b) Basra (c) Najaf (d) Erbil



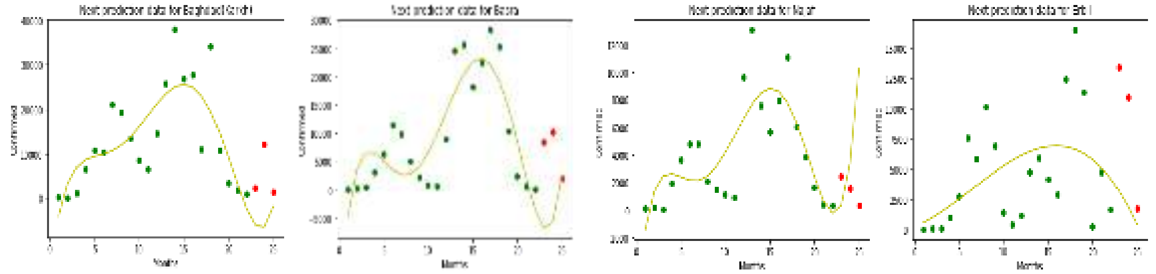
(a) (b) (c) (d)

Figure 8: fit using logistic Features (a) Baghdad (Karkh) (b) Basra (c) Najaf (d) Erbil function



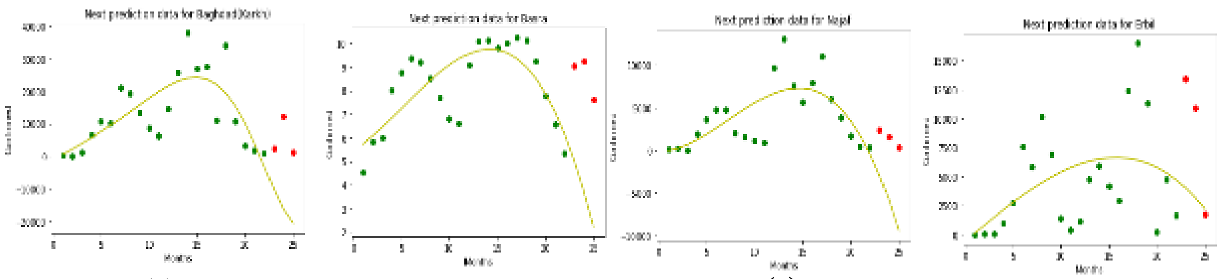
(a) (b) (c) (d)

Figure 9: fit using Polynomial function (a) Baghdad (Karkh) (b) Basra (c) Najaf (d)



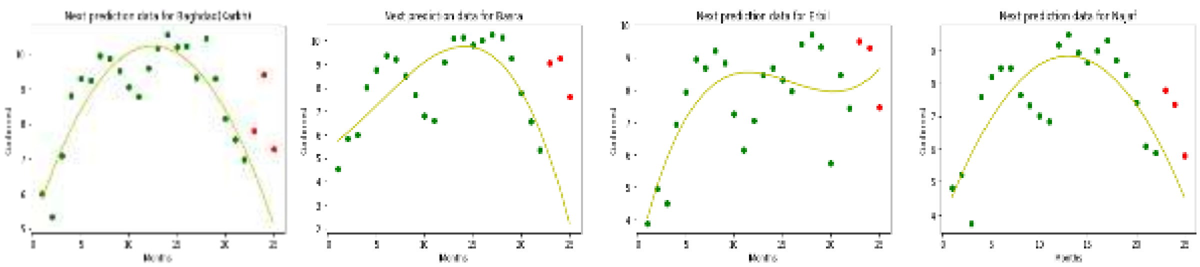
(a) (b) (c) (d)

Figure 10: fit using Polynomial Features (a) Baghdad (Karkh) (b) Basra (c) Najaf (d) Erbil



(a) (b) (c) (d)

Figure 11: fit using LassoLars function (a) Baghdad (Karkh) (b) Basra (c) Najaf (d) Erbil



(a) (b) (c) (d)

Figure 12 : Scale the input data by a log function (a) Baghdad (Karkh) (b) Basra (c) Najaf (d) Erbil

The performance of the analysis method and all results in Table 1 are shown in Figure [7], which illustrates the comparative results for four provinces after fitting the model using linear and logistic features (Figure [8]). While Figure 9 enhances the forecast using a polynomial function, and Figure 10 is fitted using polynomial features, the results were computed using the Lasso-Lars function in Figure 11, and in Figure 12, the input data by a log function for all four provinces was scaled.

5. Conclusion

It was fascinating to notice the high association between mortality rates and immunizations, as well as illness rates and vaccinations. Any doubts about vaccination efficacy were dispelled by these analytics results, and anyone who has chosen not to vaccinate can be argued against.

The study verifies that immunizations greatly improved the situation of the pandemic by leading to a decreased death and infection rate by taking a step back and assessing the whole pandemic thus far. Therefore, with new varieties of COVID-19 on the horizon, it is becoming increasingly crucial for people to grasp the significance of vaccination.

In this study, an optimization polynomial regression is presented to analyze and predict the spread of novel COVID-19. Our project encompasses different operational stages such as preparing datasets, feature extraction to forecast, and polynomial algorithm-based hyperparameter optimization. However, it assists in the optimal selection of hyperparameter values for the model. To demonstrate the superiority of the suggested model, a wide range of experimental analyses were conducted on the benchmark test dataset. The outcomes were evaluated in numerous ways. The simulation results show that the strategy performs better than the previous approaches. Furthermore, based on the data acquired thus far, we would like to employ more advanced machine learning algorithms to show a future forecast of the death/infection rate.

References

[1] B. Mohan and N. Vinod, “COVID-19: An Insight into SARS-CoV2 Pandemic Originated at Wuhan City in Hubei Province of China,” *Journal of Infectious Diseases and Epidemiology*, vol. 6, no. 4, 2020, doi: 10.23937/2474-3658/1510146.

[2] R. G. Ali Al-Suhail and L. F. Ali, “Statistical analysis of COVID-19 pandemic across the provinces of Iraq,” *Iraqi Journal of Science*, vol. 62, no. 3, pp. 811–824, 2021, doi: 10.24996/ijs.2021.62.3.12.

[3] A. A. Mohsen, H. F. AL-Husseiny, K. Hattaf, and B. Boulfoul, “A mathematical model for the dynamics of COVID-19 pandemic involving the infective immigrants,” *Iraqi Journal of Science*, vol. 62, no. 1, pp. 295–307, 2021, doi: 10.24996/ijs.2021.62.1.28.

[4] M. A. Khan, R. Khan, F. Algarni, I. Kumar, A. Choudhary, and A. Srivastava, "Performance evaluation of regression models for COVID-19: A statistical and predictive perspective," *Ain Shams Engineering Journal*, vol. 13, no. 2, 2022, doi: 10.1016/j.asej.2021.08.016.

[5] M. A. A. Al-Qaness, A. A. Ewees, H. Fan, and M. A. el Aziz, "Optimization method for forecasting confirmed cases of COVID-19 in China," *J Clin Med*, vol. 9, no. 3, 2020, doi: 10.3390/jcm9030674.

[6] T. W.-C. J. M. on C. D. 2019 and W. H. O. (WHO), "Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19)," *The WHO-China Joint Mission on Coronavirus Disease 2019*, vol. 1, no. 16-24 February, pp. 1–40, 2020, [Online]. Available: https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf

[7] M. H. R. Khan and A. Hossain, "Machine Learning Approaches Reveal That the Number of Tests Do Not Matter to the Prediction of Global Confirmed COVID-19 Cases," *Front Artif Intell*, vol. 3, no. November, pp. 1–10, 2020, doi: 10.3389/frai.2020.561801.

[8] C. Menni *et al.*, "Real-time tracking of self-reported symptoms to predict potential COVID-19," *Nat Med*, vol. 26, no. 7, pp. 1037–1040, 2020, doi: 10.1038/s41591-020-0916-2.

[9] E. Fayyoumi, S. Idwan, and H. Aboshindi, "Machine learning and statistical modelling for prediction of Novel COVID-19 patients case study: Jordan," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 5, pp. 122–126, 2020, doi: 10.14569/IJACSA.2020.0110518.

[10] 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, 2021, doi: 10.1007/s42979-020-00394-7.

[11] A. Sinha and M. Rathi, "COVID-19 prediction using AI analytics for South Korea," *Applied Intelligence*, vol. 51, no. 12, pp. 8579–8597, 2021, doi: 10.1007/s10489-021-02352-z.

[12] D. Tiwari, B. S. Bhati, F. Al-Turjman, and B. Nagpal, "Pandemic coronavirus disease (Covid-19): World effects analysis and prediction using machine-learning techniques," *Expert Syst*, vol. 39, no. 3, pp. 1–20, 2022, doi: 10.1111/exsy.12714.

[13] A. K. Gupta, V. Singh, P. Mathur, and C. M. Travieso-Gonzalez, "Prediction of COVID-19 pandemic measuring criteria using support vector machine, prophet and linear regression models in Indian scenario," *Journal of Interdisciplinary Mathematics*, vol. 24, no. 1, pp. 89–108, 2021, doi: 10.1080/09720502.2020.1833458.

[14] S. Rath, A. Tripathy, and A. R. Tripathy, "Prediction of new active cases of coronavirus disease (COVID-19) pandemic using multiple linear regression model," *Diabetes and Metabolic Syndrome: Clinical Research and Reviews*, vol. 14, no. 5, pp. 1467–1474, 2020, doi: 10.1016/j.dsx.2020.07.045.

[15] D. A. Mohammed, H. M. Tawfeeq, K. M. Ali, and H. M. Rostam, "Analysis and Prediction of COVID-19 Outbreak by a Numerical Modelling," *Iraqi Journal of Science*, vol. 62, no. 5, pp. 1452–1459, 2021, doi: 10.24996/ijs.2021.62.5.8.

[16] "وزارة الصحة العراقية." https://moh.gov.iq/ (accessed Jan. 25, 2023).

[17] H. Theil, "A rank-invariant method of linear and polynomial regression analysis I and II," 1950.

[18] A. Schneider, G. Hommel, and M. Blettner, "Lineare regressionsanalyse - Teil 14 der serie zur bewertung wissenschaftlicher publikationen," *Deutsches Arzteblatt*, vol. 107, no. 44. pp. 776–782, Nov. 05, 2010. doi: 10.3238/arztebl.2010.0776.