Quest Journals
Journal of Software Engineering and Simulation
Volume 8 ~ Issue 5 (2022) pp: 09-15
ISSN(Online) :2321-3795 ISSN (Print):2321-3809
www.questjournals.org

Research Paper

A Model for Forecasting Covid-19 Cases in Nigeria Based on Time Series Data Using Support Vector Machine

[#1] Chizi Aloy-Okwelle
(Computer Science Department, Rivers State University Nigeria)
[#2] Dr Daniel Mathias
(Computer Science Department, Rivers State University Nigeria)
[#3] Dr V.I.E Anireh
(Computer Science Department, Rivers State University Nigeria)
Corresponding Author: Chizi Aloy-Okwelle

Abstract: *This study is focused on forecasting the spread of Covid-19 cases in Nigeria. It's critical to have early warning systems in place that can predict how much the disease will affect society, allowing the government to organize resources, set policies, and provide survivors with immunity passports while still using the same plasma for treatment. The objective of this work is to produce a real-time forecast using the Support Vector Machine. Data on the daily spread of Covid-19 from February 27, 2020 to November 2, 2021 where collected from the NCDC database repository. The data is in the form of a .csv file updated on regular basis. Pandas library in python was used for loading, preparing, and analyzing the dataset and then applied to the Support Vector Machine. The training and testing process of the dataset was carried out using Jupyter Notebook in python. The result of the model shows that there will be an increase in covid-19 case in the next two months ranging from 2000 to 4000, and the model shows an accurate forecasted measure of 97.6. Through the President's Task Force, the federal government of Nigeria may make more informed decisions about additional measures to contain the spread of the virus, using the forecasted pattern. The model may also be used to investigate the efficiency of the lockdown in preventing the spread of Covid-19 in Nigeria.*
Keywords- *COVID-19, Support Vector Machine, Machine learning.*

***R**eceived 01 May, 2022; **R**evised 10 May, 2022; **A**ccepted 12 May, 2022 © The author(s) 2022.*
P**ublished with open access at **www.questjournals.org

I. INTRODUCTION

The novel Coronavirus (COVID-19) is an inflammation disease that appeared first in China (Wuhan city) in December 2019. The disease causes respiratory ailments, the clinical characteristics of Covid-19 include, fever, cough, dyspnea and viral pneumonia in server cases [1]. The main problem of these symptoms is that there is virus infected asymptomatic patients. The virus has spread rapidly around the world and has become a global pandemic [2]. The effects have been devastating on daily lives, public health and the global economy. The virus spreads quickly between people and provoked the closure of boarders in many countries and the confinement of millions of citizens to their homes due to infected people, which amounts to 155,506,494 confirmed cases worldwide at time of writing this paper in April 2020 [3]. Compared to other populations, the virus has a greater impact on people older than 65 and younger than 10, and it may also depend on some type of personal health history [1].

On February 27, 2020, Nigeria announced the first COVID-19 verified case, when an Italian person tested positive for the virus. As a result, the government, like other governments in afflicted nations, adopted non-pharmaceutical efforts to halt the virus's spread. The Nigerian government-imposed travel restrictions on 15 high-incidence nations, including China, Italy, Iran, Norway, South Korea, Spain, Japan, France, Germany, the United States of America, the United Kingdom, the Netherlands, Switzerland, Sweden, and Austria, as of March 20, 2020 [5]. On March 20, 2020, schools were ordered to close for a period of time. On March 30,2020, a number of measures were adopted, including border restrictions, work-from-home arrangements for government workers, limitations on social and religious gatherings, the wearing of facial masks, social distance, states and national lockdowns.

The number of cases in the country has continued to rise, and they have now spread throughout many states. As of 20th June, 2021, 167155 cases have been confirmed, 163540 cases have been discharged and 2117 deaths have been recorded in 36 states and the Federal Capital Territory [6]. Gradually the mortality rate is increasing and it's an alarming factor for the whole world. The number of cases is growing nationwide in Nigeria and local transmission has surged relatively high.

One of the fundamental problems facing the medical research community is that the properties of the virus is changing and new variants have emerged since the virus was introduced into the human body in China. This non-monotonic and uncertain behavior of this virus spread makes it different and more dangerous from other viruses making it very difficult to monitor and predict.

The case fatality rate in Nigeria is increasing [7]. Nigeria's anticipated CFR level of 0.03 percent will be maintained in the foreseeable future, if there are no significant changes in healthcare facilities, detection technologies, continuing clinical treatments, or other factors.

The prediction of COVID-19 transmission in real time is critical since a worldwide pandemic is on the horizon. The purpose of this study is to forecast the spread of the Covid-19 cases in Nigeria. This can provide important information for responding to outbreaks and understanding the impact of government strategies to contain the epidemic of the disease. In addition, the prediction can help in planning resources, setting government policies, equipping survivors with immune passports, and using the same plasma for care.

II. RELATED WORKS

Prediction and forecasting strategies include a variety of methods. ML methods used for prediction include support vector machines, linear regression, logistic regression, naive bays, decision trees (random forest and ETC), nearest neighbor methods, and neural networks (multilayer perceptron), etc. This section covers some existing literature publications as well as research contributions.

[8] trained support vector regression, step-down linear regression, gradient boosted regression tree model, and negative binomial regression models. Weekly dengue cases, climatic variables and Baidu search queries were utilized as input characteristics to forecast infected cases in six Chinese provinces. RMSE was used to evaluate and validate the performance of all models across provinces and forecasting windows. However, the study did not take into account the interdependence of data from different provinces.

[9] proposed a model focused on the analysis of the spread of Covid-19 in Nigeria, applying statistical models and using data from the NCDC. They collected a daily spread of data from 27 February to 26 April 2020. Using the R software, data was collected to construct the autoregressive integrated moving average (ARIMA) model. They carried out Stability analysis and stationarity test, parameter test, and model diagnostic test. Finally, the ARIMA model was assessed using the AICc model selection criteria for fitting, selection, and prediction accuracy. The model produced a ten-day forecast, which shows a significant increasing trend in the spread of the COVID-19 in Nigeria over the given time period.

[10] created an Auto-Regressive Integrated Moving Average (ARIMA) model to forecast the epidemiological trajectory of COVID-19 incidence in Italy, Spain, and France, the three most impacted European nations. ARIMA models were created to forecast the epidemiological trend of COVID-19 incidence in Italy, Spain, and France, the three most afflicted European nations. The World Health Organization website was used to gather COVID-19 prevalence data from February 21, 2020 to April 15, 2020. Several ARIMA models with various ARIMA parameters were developed. The study demonstrates that ARIMA models are capable of forecasting COVID-19 prevalence in the future.

[11] used regression models for forecasting. The SEIR model and the regression model were used to make predictions based on data obtained from the John Hopkins University repository between the 30th of January and the 30th of March 2020. RMSLE was used to assess the models' performance, and the SEIR model received a score of 1.52 while the regression model received a score of 1.75. Between the SEIR and Regression models, the RMSLE error rate was determined to be 2.01.

[12] proposed a method for forecasting future cases based on existing data. They used machine learning approaches for predicting the chance of being infected with Covid-19 and also for forecasting the number of positive cases. They employed the autoregressive integrated moving average time series to forecast confirmed cases for various Indian states. Random forest and extra tree classifiers were utilized, both of which have an accuracy of more than 90%. The extra tree classifier is the more accurate of the two, with a score of 93.62 percent.

[13] analyzed time series data using SVM. They collected worldwide data from 22 January 2020 to 25 April 2020, including regular confirms, deceased cases and recovered cases from Johns Hopkins University Public Repository Center for Systems Science and Engineering (CSSE). The SVM learning algorithm generated optimum hyperplane which separate the data accordingly.

III. METHODOLOGY

An experiment was carried out in three phases. The first phase comprises of the data collection, data cleaning and feature extraction. The second stage comprises of analysis and visualization of the covid-19 data, and finally the third phase comprises of the model training, and the analysis of the predicted results through graphs and charts.

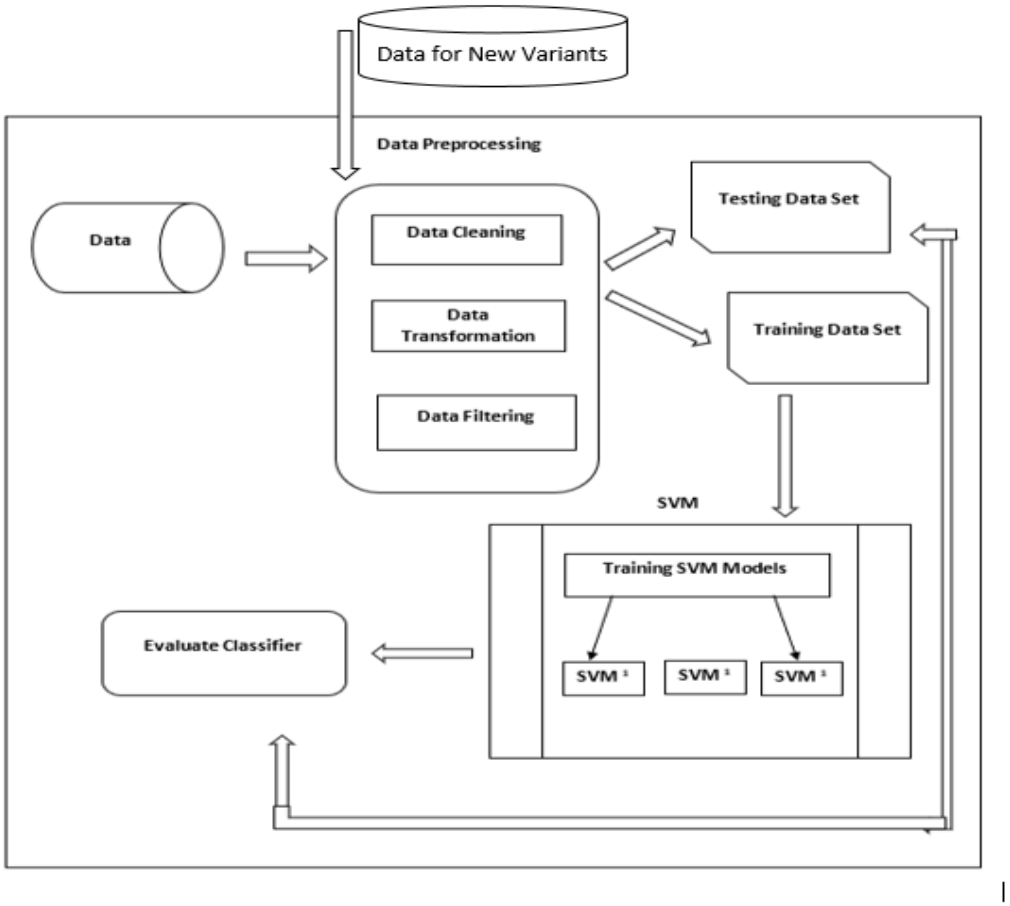

Figure 1: Architecture of the Proposed System

A. Data Collection

Data comprising confirmed, recovered, and death cases as variables; were downloaded as a .csv file from the Nigeria Centre for Diseases Control (NCDC) official COVID-19 site https://covid19.ncdc.gov.ng/ . The dataset properties comprise of different columns like State Affected, Total confirmed case, weekly confirmed cases, Total Recoveries, Total Deaths and Total Test conducted. The covid-19 dataset used here comprises of the whole 36 states in Nigeria from February 27, 2020 to November 2, 2021. The information is presented in a cumulative format. We computed the different time series from the cumulative dataset to acquire the values based on daily new case basis. The covid-19 data can be seen in figure 2.

	State	Total_confirmed_cases	Last_week_c	Total_Recoveries	Last_week_r	Total_Deaths	Last_week_d	Active_cases	Total_test	Last_week_t
0	Abia	2019	6	1989	43	30	0	0	34820	161
1	Adamawa	1157	0	1098	0	32	0	27	27990	201
2	Akwa Ibom	4348	0	4076	0	44	0	228	43023	397
3	Anambra	2387	18	2333	0	19	0	35	43285	202
4	Bauchi	1703	26	1633	24	19	1	51	34074	432
5	Bayelsa	1240	8	1199	11	28	0	13	29531	510
6	Benue	1858	73	1455	0	25	0	378	43093	267
7	Borno	1356	0	1306	0	38	0	12	23616	125
8	Cross River	634	9	597	8	25	0	12	15863	150
9	Delta	4142	554	2556	0	110	0	1476	66307	1001
10	Ebonyi	2062	3	2004	1	32	0	26	34981	143
11	Edo	6591	23	6318	154	224	0	49	98979	1576
12	Ekiti	1774	12	1718	14	28	0	28	44747	861
13	Enugud	2761	35	2609	20	29	0	123	56988	823
14	FCTd	23261	144	20984	83	220	3	2057	451655	9896

Figure 2: Covid-19 dataset

	Name of State	Number of Confirmed Cases	Number of Deaths	Number of Recoveries	Number of Active Cases	Mortality Rate
0	Lagosd	77808	752	76918	138	0.009665
1	FCTd	23261	220	20984	2057	0.009458
2	Rivers	12695	154	12481	60	0.012131
3	Kaduna	10026	79	9870	77	0.007880
4	Plateau	9766	70	9527	169	0.007168
5	Oyo	8754	191	8543	20	0.021819
6	Edo	6591	224	6318	49	0.033986
7	Ogun	5373	80	5291	2	0.014889
8	Ondo	4562	98	4360	104	0.021482
9	Kano	4354	116	4162	76	0.026642
10	Akwa Ibom	4348	44	4076	228	0.010120
11	Delta	4142	110	2556	1476	0.026557
12	Kwara	3952	63	3650	239	0.015941
13	Osun	3006	86	2877	43	0.028609
14	Enugud	2761	29	2609	123	0.010503
15	Gombe	2582	57	2501	24	0.022076
16	Nasarawa	2503	39	2345	119	0.015581
17	Anambra	2387	19	2333	35	0.007960
18	Katsina	2226	35	2185	6	0.015723
19	Imod	2078	41	1920	117	0.019731
20	Ebonyi	2062	32	2004	26	0.015519
21	Abia	2019	30	1989	0	0.014859
22	Benue	1858	25	1455	378	0.013455
23	Ekiti	1774	28	1718	28	0.015784
24	Bauchi	1703	19	1633	51	0.011157
25	Borno	1356	38	1306	12	0.028024
26	Taraba	1242	30	1186	26	0.024155
27	Bayelsa	1240	28	1199	13	0.022581
28	Adamawa	1157	32	1098	27	0.027658
29	Niger	1057	20	998	39	0.018921
30	Sokoto	807	28	777	2	0.034696
31	Cross River	634	25	597	12	0.039432
32	Jigawa	604	17	585	2	0.028146
33	Yobe	502	9	490	3	0.017928
34	Kebbi	464	16	433	15	0.034483
35	Zamfara	302	8	268	26	0.026490
36	Kogi	5	2	3	0	0.400000

Figure 3: New trainable data

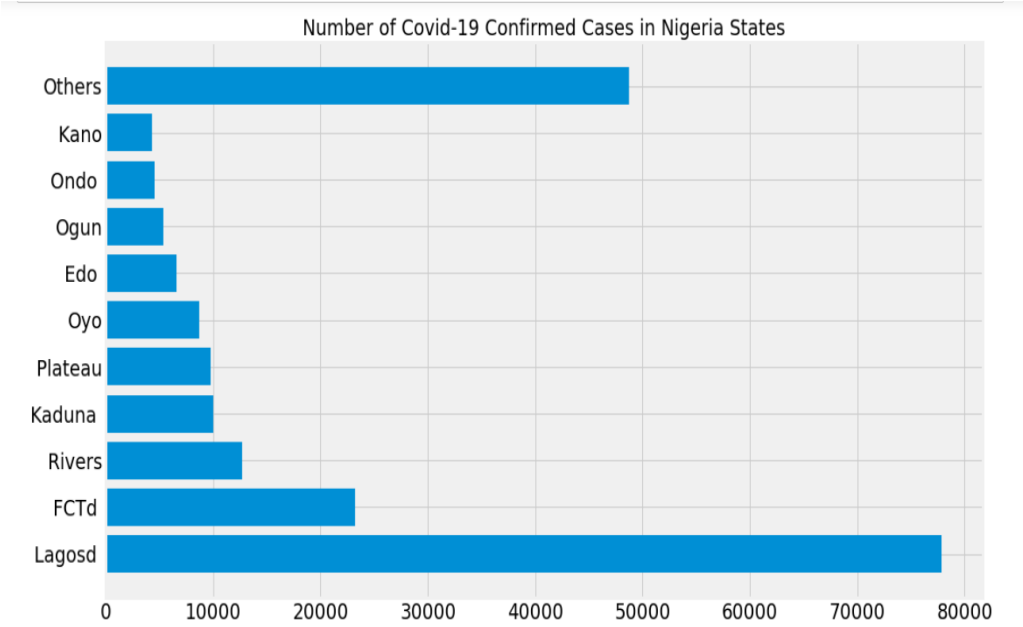

Figure 3.1: Number of confirmed cases in Nigeria States

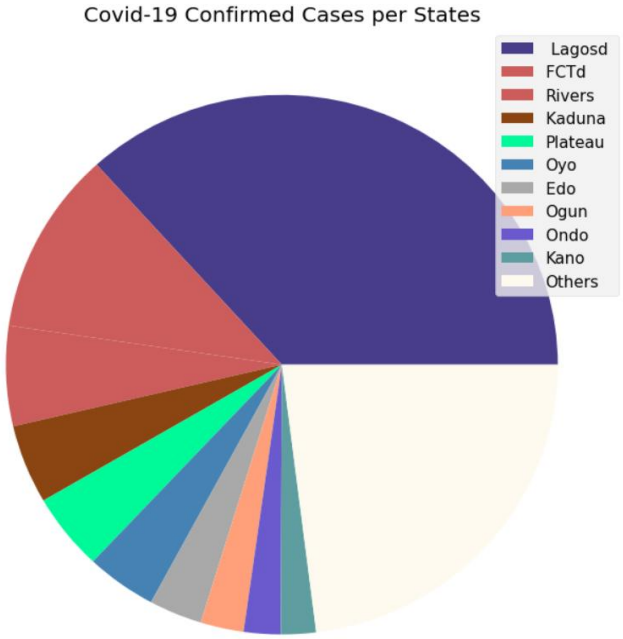

Figure 4: Pie Chart representation of covid-19 cases

B. Data Pre-processing

The data was further pre-processed and cleaned by checking for Nan values and the removal of comma (,) from the covid-19 data. The dataset was further processed by appending the mortality rate to it.

C. Model Training/Future Forecast

Machine learning time series considers the time parameter when evaluating other inputs**.** Support vector machine algorithms was applied in training a model for predicting the trend of covid-19 in Nigeria. The model was trained on the covid-19 data that comprises of the total number of covid-19 cases since the year 2020 to 2nd of November 2021. The dataset is split for Training (80%) and Test (20%) using train_test_split() function imported from class model_selection of sklearn python library. The variables for training and testing are kept for future analysis.

IV. RESULT & DISCUSSION

COVID-19 pandemic cases were forecasted by the SVM machine learning algorithm. The results show that the model performed well. The overall distribution data of COVID-19 from February 27, 2020 to November 2021, were collected, trained and analyzed. Day by day, the incidence increased in a wave-like pattern. When the Federal Government of Nigeria enforced the first lockdown on March 27, 2020, there was a trend of rectilinear rise afterwards.

Figure 3 shows the covid-19 cases used as data. The columns are made up of ten (10) columns, starting from state column to the last week test column. The state column shows states that are affected by covid-19, the total_confirmed column shows the total number of covid-19 cases in each of the 36 states. The last_week_c shows the total number of covid_19 cases gotten in a week, Total_Recoveries column shows the total number of recovered cases in each of the affected states, the last_week_r shows the weekly recoveries, Total_Deaths columns shows the total number of covid-19 deaths, the last_week_d shows the total number of deaths that occurs weekly because of covid-10, the Active cases column shows the total number of active cases, the Total_test column shows the total number of covid test, and finally the last_week_t shows the weekly test conducted.

Figure 5 shows the predicted result of Support Vector Machine. The result shows that there will be an increase of covid-19 cases from 2000 to 4000 if the current rate of daily new cases prevails. The tested data (blue line) shows that the covid-19 cases will be increased over 12,000 in the next two months. The tested data is 20% of the training data.

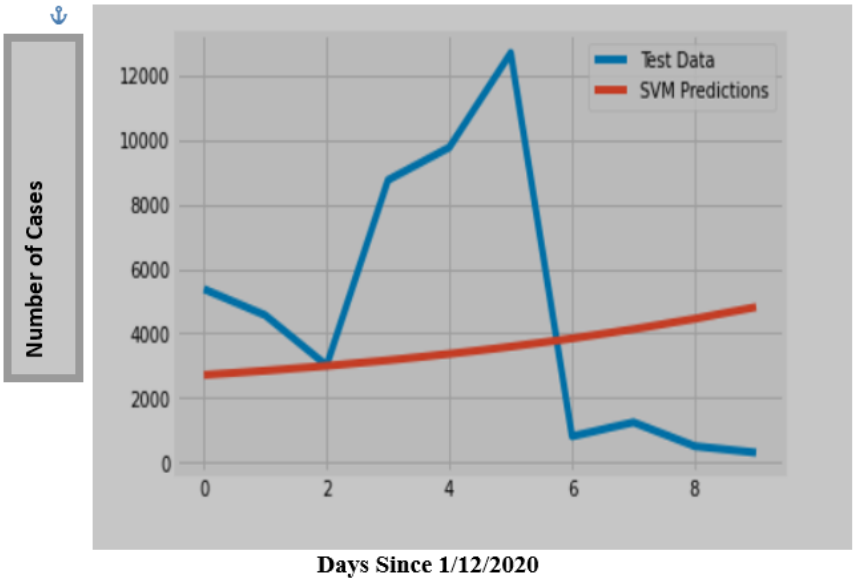

Figure 5: Result of Support Vector Machine

V. CONCLUSION

This paper can be used by different researchers to learn how ML can be employed to forecast not only this situation but also other cases. The model gives results on the basis of data developed by information gotten from the NCDC database repository. SVM model is one of the good quality machine learning technique used for prediction because of its simplicity. This model predicts that the number of cases in Nigeria will increase in coming weeks. There is a need for effective strategies to prevent and monitor disease spread. To maintain the national economy's growth, the government must carry out required resource allocation of medical supplies, healthcare equipment, agricultural, and manufacturing activity control. As a result, developing an effective forecasting model that can provide the government an advantage in deciding on macroeconomic emergency plans is critical

Furthermore, the performance of the model can be improved or the model can give more accurate data if more datasets are available.

REFERENCES

[1]. Li, J. & Fan, J. G. (2020). Characteristics and mechanism of liver injury in 2019 coronavirus disease. *Journal of Clinical and Translational Hepatology*, 8(1), 13.

[2]. Wu, F., Zhao, S., Yu, B., Chen, Y. M., Wang, W., Song, Z. G. & Zhang, Y. Z. (2020). A new coronavirus associated with human respiratory disease in China. *Nature*, 579(7798), 265-269.

[3]. World Health Organization (2021), WHO Coronavirus (COVID-19) Dashboard, https://covid19.who.int/. (accessed 21 April 2021)

[4]. Nejo, A. (2020). Coronavirus: Work from home, Fayemi tells Ekiti civil servant, The Punch newspaper https://punchng.com/breaking-coronavirus-work-from-home-fayemi-tells-ekiti-civil-servants (2020) (accessed 21 March 2020)

[5]. NCDC (2021) COVID-19 NIGERIA. https://covid19.ncdc.gov.ng/

[6]. Suleiman, A. A., Suleiman, A., Abdullahi, U. A. & Suleiman, S. A. (2021). Estimation of the case fatality rate of COVID-19 epidemiological data in Nigeria using statistical regression analysis. *Biosafety and Health*, *3*(1), 4-7.

[7]. Guo, P., Liu, T., Zhang, Q., Wang, L., Xiao, J., Zhang, Q., ... & Ma, W. (2017). Developing a dengue forecast model using machine learning: A case study in China. PLoS neglected tropical diseases, 11(10), e0005973

[8]. Ibrahim, R. R. & Oladipo, O. H. (2020). Forecasting the spread of covid-19 in Nigeria using box-jenkins modeling procedure. medRxiv.

[9]. Ceylan, Z. (2020). Estimation of COVID-19 prevalence in Italy, Spain, and France. *Science of The Total Environment*, 729, 138817.

[10]. Pandey, G., Chaudhary, P., Gupta, R. & Pal, S. (2020). SEIR and Regression Model based COVID-19 outbreak predictions in India. arXiv preprint arXiv:2004.00958.

[11]. Painuli, D., Mishra, D., Bhardwaj, S., & Aggarwal, M. (2021). Forecast and prediction of COVID-19 using machine learning. In Data Science for COVID-19 (pp. 381-397). Academic Press.

[12]. Singh, V., Poonia, R. C., Kumar, S., Dass, P., Agarwal, P., Bhatnagar, V., & Raja, L. (2020). Prediction of COVID-19 corona virus pandemic based on time series data using Support Vector Machine. Journal of Discrete Mathematical Sciences and Cryptography, 23(8), 1583-1597.