International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 11, No.1 (Mar-2022)

https://dx.doi.org/10.12785/ijcds/110178

Selecting a Better Classifier Using Machine Learning for COVID-19

Muhammad Imran[1], Mian Usman Sattar[2], Hamza Wazir Khan[3], Abdul Ghaffar[4] and Dr. Hammad Mushtaq[5]

[1,3,,5]*Department of Information Systems, University of Management and Technology, Lahore, Pakistan.*
[2]*Department of Management, School of Business, Beaconhouse National University, Lahore, Pakistan*
[4]*Department of Management, University of Management and Technology, Lahore, Pakistan.*

E-mail address: statisticsgcu@gmail.com, usman.sattar@bnu.edu.pk, hamza.khan@umt.edu.pk, abdul.ghaffar@umt.edu.pk, hammad.mushtaq@umt.edu.pk

Received 11 Aug. 2020, Revised 21 Dec. 2021, Accepted 21 Jan. 2022, Published 31 Mar. 2022

Abstract: This paper will elaborate that how timely available data and Machine learning algorithms can help in determining premature exposure of coronavirus (COVID-19) and aided the world in formulating to reduce the loss. We will investigate which machine learning algorithms are best fit to predict COVID-19 data sets. In this study our focus will be on the spread of COVID-19 internationally in different countries. This study will serve as a resource for the future research and development on COVID-19 by producing better research in this field. To achieve the outcomes and future forecasting of COVID-19, we analyze the records and datasets of COVID-19 through Machine Learning algorithms. For this purpose, we used six algorithms to construct classifiers such as K-Nearest Neighbor (K-NN), Decision Tree, Support Vector Machine (SVM), Naïve Bayes, Logistic Regression and Random Forecast. These algorithms were applied on Python a machine learning software. The dataset is acquired by WHO data sets and data sets provided online at GitHub and compiled and organized by different communities to track the spread of the virus. The Performance of the best classifier will be measured using Accuracy. The model developed with Decision Tree is one of the most efficient classifier with the highest percentage of accuracy of 99.85 percent, and is followed by Random Forecast with 99.60 percent, Naïve Bayes with 97.52 percent accuracy, Logistic Regression with 97.49 percent accuracy, Support Vector Machine with 98.85 percent accuracy and K-NN with 98.06 percent accuracy. In our research, we discussed two types of classification: Binary and Multinomial. Support Vector Machine and Decision Tree give us precise results. Other classifier models gave satisfactory outcomes. The outcomes may be helping to predict the future circumstances of COVID-19. From the past studies we have used Autoregressive integrated moving average (ARIMA) model for time series data. SIR models to check the spread of Nowcasting and forecasting the spread of 2019-nCoV in China and worldwide.

Keywords: COVID-19; MERS;SARS; SVM;K-NN

1. Introduction

In December 2019, some unknown pneumonia diseases and cases emerged from Wuhan, China. It was reflected that a cluster of patients who visited the Chinese wholesale Seafood market was the source of the disease. Later, it was diagnosed that pangolins and snakes at the seafood market were likely to be the intermediate host of the coronavirus [1]. An unusual coronavirus was discovered as a mature virus tentatively. According to its physical and biological characteristics such as Taxonomy and phylogeny, it is identified as a sister of severe acute respiratory syndrome (SARS) and known as advanced SARS-CoV-2 today [2]. COVID-19 has reached almost 212 countries and territories as of May 4, 2020, and triggered the global health emergency and affecting about 3,593,627 around the globe which are reported cases with 1,166,079 recoveries and 249,112 total deaths [3]. This paper will elaborate that how timely available data and Machine learning algorithms can help in early determine the premature coronavirus (COVID-19) and aided the world in formulating to reduce the loss. It was on January 9th, 2020, when WHO issued its first official warning on the spread of coronavirus after receiving reports regarding the coronavirus from Chinese officials [4]. Viral outbreaks over the last two decades, we will explore in this paper that how early and timely detection of disease can reduce the time as computation technology has improved [5]. Just we see in the case of (COVID-19) where disease detection was done in a short period. If we compare it with the past epidemic severe acute respiratory syndrome (SARS) where it took almost four months [6] Previously,

E-mail address: statisticsgcu@gmail.com, usman.sattar@bnu.edu.pk, hamza.khan@umt.edu.pk, abdul.ghaffar@umt.edu.pk, hammad.mushtaq@umt.edu.pk

a lot of research has been done on Coronavirus but we will introduce how we can add to this by using Machine Learning Algorithm which has not been used before. It was the need of the hour as this virus is affecting many countries rapidly. How we can make a better prediction by using Machine Learning Algorithms and reduce the risk of spread of this virus where governments can take appropriate steps to tackle the pandemic. By using our research most of the government agencies could benefit from it. It will be also useful for academics to learn more about this topic. has reached almost 212 countries and territories as of May 4, 2020, and triggered the global health emergency and affecting about 3,593,627 around the globe which are reported cases with 1,166,079 recoveries and 249,112 total deaths [3]. This paper will elaborate that how timely available data and Machine learning algorithms can help in early determine the premature coronavirus (COVID-19) and aided the world in formulating to reduce the loss. It was on January 9th, 2020, when WHO issued its first official warning on the spread of coronavirus after receiving reports regarding the coronavirus from the Chinese officials [4]. Viral outbreaks over the last two decades, we will explore in this paper that how early and timely detection of disease can reduce the time as computation technology has improved [5]. Just we see in the case of coronavirus (COVID-19) where disease detection was done in a short period. If we compare it with the past epidemic severe acute respiratory syndrome (SARS) where it took almost four months [7] Previously, a lot of research has been done on Coronavirus but we will introduce how we can add to this by using Machine Learning Algorithm which has not been used before. It was the need of the hour as this virus is affecting many countries rapidly. How we can make a better prediction by using Machine Learning Algorithms and reduce the risk of spread of this virus where governments can take appropriate steps to tackle the pandemic. By using our research most of the government agencies could benefit from it. It will be also useful for academics to learn more about this topic.

2. Related Work

Like as we all know that as just months back in December 2019 Coronavirus has been detected in China. So, there is not enough research and clinical trials have been performed. Researchers are in the initial stages of finding out the cure for COVID-19. But we thank the internet where a lot of scholarly articles are available where we can find out how academics have put all their efforts to find out the most valuable information so we can get rid of this pandemic. This virus is considered by most researchers as the sister of SARS-Cov which was also started in China some 17 years back in November 2002. At that time, it took China almost 4 months to detect the SARS-Cov and they could not handle that situation well. Now, as compared with the SARS period, the capacity to handle COVID-19 has tremendously improved and this time it was COVID-19 was diagnosed in just a couple of days. At the same time, the extraordinary clinical experiments and level of evidence or indication will provide valuable evidence-based data for clinical diagnosis and identification for the treatment of COVID-19 [7]. The current outburst of COVID-19 fascinates many researchers to assist and explore a trail to recover. In order to identify COVID-19, [8] provide a paradigm for utilizing mobile phones while quarantined. [9] created a predictive algorithm to distinguish suspicious individuals before they become critically ill [10] used a regression model to investigate the rapid evolution of COVID-19.

It is believed by many researchers that like the previous Like as we all know that as just months back in December 2019 Coronavirus has been detected in China. So, there is not enough research and clinical trials have been performed. Researchers are in the initial stages of finding out the cure for COVID-19. But we thank the internet where a lot of scholarly articles are available where we can find out how academics have put all their efforts to find out the most valuable information so we can get rid of this pandemic. This virus is considered by most researchers as the sister of SARS-Cov which was also started in China some 17 years back in November 2002. At that time, it took China almost 4 months to detect the SARS-Cov and they could not handle that situation well. Now, as compared with the SARS period, the capacity to handle COVID-19 has tremendously improved and this time it was COVID-19 [7]

diagnosed in just a couple of days. At the same time, the extraordinary clinical experiments and level of evidence or indication will provide valuable evidence-based data for clinical diagnosis and identification for the treatment of COVID-19. The current outburst of COVID-19 fascinates many researchers to assist and explore a trail to recover. For the detection of COVID-19 [8] present a model of using mobile phones during the quarantine [9] developed a predictive model to distinguish the initial screening of suspected patients before they are critically ill [10] established a regression model to analyze the drastic development of COVID-19. It is believed by many researchers that like the previous types of corona viruses as Middle East Respiratory Syndrome (MERS)and SARS, COVID-19 is also started from the bats which are believed to be the natural host of COVID-19 as bats have extensive geographical displacement capability of flying. A various type of coronavirus has been detected from the different species of bats throughout Asia, America, Europe, and Africa. Bats affected by these viruses sometimes show clinical symptoms of illness [11]. Various types of rhinophid bats in China carry out inherently different SARS-sort of coronavirus, out of which some belong to SARS-CoV and can directly transfer in humans [12]. Before COVID-19, the world was already facing this type of virus-like Middle East Respiratory Syndrome (MERS). This infectious disease outbreak around the globe. Coronavirus has a large family, and all its spices severely affected living organisms especially humans and animals. SARS, MERS, and Corona are renowned and common types [13]. At the start of the Middle East respiratory

syndrome (MERS), it spread into other species as COVID-19 did. Therefore, these diseases are known as sisters of each other as they have the same symptoms and host. The Early diagnosis of the disease was the need of the hour. Academics applied machine learning and data mining models to early predict and diagnose large and complex datasets of health care where they can support to control the outbreak of the virus in other countries. The data mining model could be used on a complicated and large amount of data to obtain exact statistical patterns [14]. In the medical zones, the data mining model can assist to calculate and detect the infection and help in the medical decision-making process. Researchers and academics use data mining tentative, applications, techniques, and methods to find appropriate and accurate outcomes. For forecasting and improvement of recurring diseases, Data mining can apply to the records of medical history. Different data mining algorithms such as Naïve Bayes, Decision Tree, Support Vector Machine (SVM), and Artificial Neural Networks (ANN) are used. Through experimentation, SVM gives better results than the other classifier on the heart patients' dataset, and the Naïve Bayes classifier accomplishes the highest precision on the diabetes dataset [15]. Machine learning algorithms turn out to be the most effective and operative for sensitivity, precision, and accuracy. To check the four classifiers' performance, we experimented on Wisconsin Breast cancer data from the UCI Machine Learning repository: Naïve Bayes, Support Vector Machine, Decision Tree, and K-Nearest Neighbors Algorithm. For model development as, correctly classified cases, wrongly classified cases and precision, the performance of all classifiers was measured. The result shows that the Support Vector Machine is the best prediction classification for the highest accuracy and highest failure rate breast cancer data. [16].

3. RESEARCH METHODOLOGY

The classifier models forecast the accuracy and precision of COVID-19 but not all classifier models need to give exact and appropriate results because of every model and statistical classifier has its qualities and drawbacks. But some are given extraordinary results. In this study, we analyze the outcomes of classifier models through binary and multiclass classification of datasets. Six algorithms were incorporated to develop classification models, which were K-NN, Decision Tree, and SVM,Naïve Bayes, Logistic Regression and Random Forecast. These algorithms are applies using Python, A Machine Learning Software. The outcomes of best classifier will be measured using Accuracy. Multiple machine learning models can be used to make the predictions by converting time series data into supervised learning by re-engineering features using the target variable and/or its co-variates. From past studies, we can use the ARIMA model for time series data. SIR models to check the spread of Nowcasting and forecasting the spread of 2019-nCoV in China and Worldwide [15], [16].

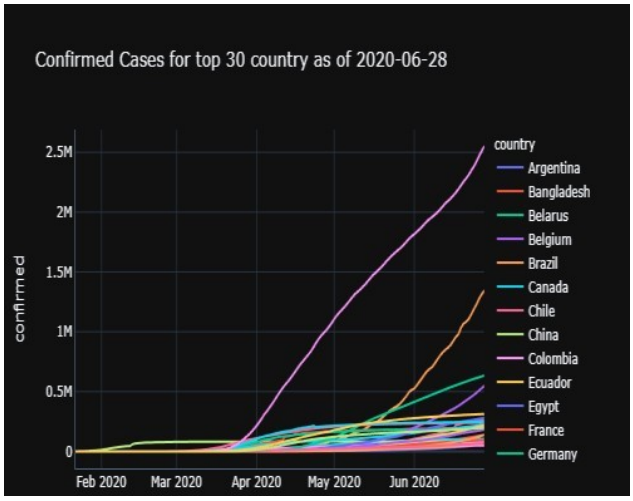

Figure 1. COVID 19 Worldwide Situation

A. Data Set Collection and Description

The dataset is acquired by WHO data sets and data sets provided online and different websites such as the Kaggle website compiled and organized by different communities to track the spread of the virus. In this research, we utilized the publicly available datasets comprised of the affected countries' daily registered cases such as affected patients, recovered patients, deaths, and a total number of performed COVID-19 tests from 22 January 2020 [5]. As this is still a developing story, the number of data is increasing exponentially with each passing day, this will help to train our models more accurately and the results can be measured and compared more accurately with the increasing period. This study makes use of two data sets. The first is an epidemiological study of COVID-19 patients in South Korea, which included 3,360 cases. 3 and 8 qualities containing the patient's ID, the international number assigned by the Korean Centre For Disease Control, Gender, Year Of Birth, City, Country, Province disease (False: No Disease / True: Have Disease), deceased date (the date of death), infection order (the ID of the infected patient), the date of onset symptom infection case, released date (the date of release).Second data set is extracted from the Kaggle with the same attributes, but this dataset is for worldwide trend [17].

Relevant and appropriate attributes and elements were taken out from the source dataset through cleaning and arranging data. We take only 1,505 cases on the bases of these elements: age, gender, infection cases, state of patients, and no. of days (number of days means the day of corona test confirmation to the day of recovery or death).We discuss only two states of patient either patient recovered or die. Missing values in the dataset create noise which affects the prediction accuracy of the algorithm, which leads to a false conclusion [17]. Therefore, we have utilized the last survey contained forward imputation technique to detect dataset's missing values.

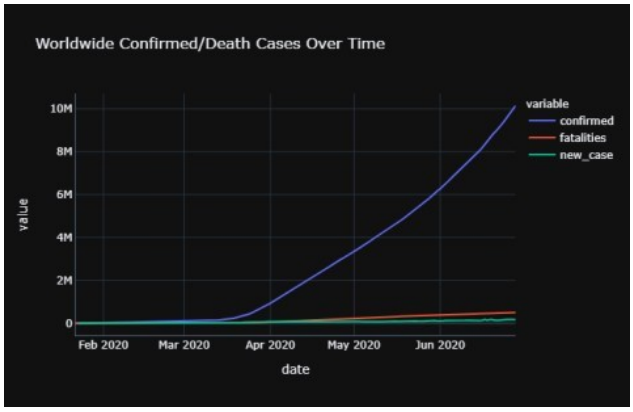

Figure 2. Confirmed cases for top 30 countries as of 2020-06-28

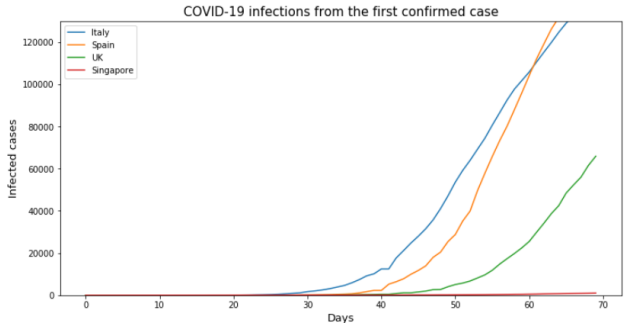

Figure 3. Comparison of infectious cases between, Spain, UK and Singapore

B. Tools

Python and its libraries along with data visualization tools such as Power BI will be used to derive the results. For making Graphs we have used Plotly software.

4. RESULTS AND PREDICTIONS

In this section, we will be analyzing our data set and making predictions of the spread of this COVID-19 internationally as per our objective of the study. Firstly, we will see the graphical presentation of this disease and making analyzing all the statistics about this topic. Also, we will be comparing different trends of deaths, recoveries between different countries and regions. Secondly, I will be comparing machine learning algorithm and will predict which machine learning are the best fitted to predict the COVID-19. Finally, we will also see prediction based on SIR and ARIMA model for the COVID-19 data set.

A. COVID- 19 Worldwide Situation

Figure 2 illustrates the international projection for COVID-19 spread. The graph shows confirmed cases, new cases and deaths within the respective periods of COVID 19. The duration of the x-axis and the number of confirmed COVID-19 instances on the y-axis are displayed. Figure 4 shows confirmed COVID-19 patients, infected cases and emerging cases as predictive evidence GitHub is the source of the global dataset on which ML model is trained for prediction. When we see at the confirmed cases world-wide it looks like that the graph is showing exponential development. The number of confirmed cases is increasing very rapidly especially at the end of June 2020 which is an alarming situation for the whole world. From the above figure 4, we can see that the total number of cases from Jan 22, 2020, to June 27, 2020, about ten million new cases have been reported.

Figure 3, depicts that as coronavirus hits at very first but fortunately its trend going to slow down since March 2020 which is a good sign but as the travel among the countries was not fully banned from China and a very wicked wave of coronavirus cases hits back at Europe in March especially countries like Spain, Italy, Germany, France, and the UK. But most critically when COVID-19 hits back in the US, the growth rate was much faster than in Europe and China. From figure 2 above we can see that its main spread was started in the middle of March which was even faster than in Italy. Figure 2 tells that at this stage the USA is the most terribly affected country in the world with coronavirus. We can also see that cases in Brazil and Russia also begin on increasing in June. From Figure 3, it is much clarified to say that Italy and Spain are the countries with rapid growth in the number of Covid19 cases following the UK with somehow low cases. The Country like Singapore is the best where coronavirus cases were flattening from the start to the end which is an ideal situation. Others should follow and take proactive measures to overcome the pandemic. From figure 3, we could also see that after one month of first case detection, cases started exponentially increasing irrespective of Singapore.

• Italy is predicting an alarming situation for COVID-19 due to 120,000 cases. The infectious curve for Italy is almost steep as more than 2 percent population proportion of the country is suffering from COVID-19, which impacts terribly for the country. • We can see that in the 2nd phase of the pandemic, Spain and Italy has the same number of infectious cases, but Spain has more than 120,000 cases. As Spain has less population but infectious cases are increased and the increasing rate is 3 percent. • The United Kingdom. UK has less no. of cases because she performed no. of tests to confront COVID-19. She has about 40,000 cases, which is 0.6 percent of the entire population. • Singapore. Singapore is quite an isolated island. She has less no. of tours and travels competitively for the other 3 countries. But still, she has less no. of cases even her general rate is increase day by day. The virus spread rapidly in its beginning, but fortunately, the infectious curve slop was not increasing from the past two weeks as only 0.2 of the population has been infected. • Coronavirus (COVID-19) Infected Patients Recovery in South Korea: A Comparison of Machine Learning Algorithm Using Accuracy. • Now we will be using a data mining algorithm for the infected patient's recovery in South Korea and will predict which

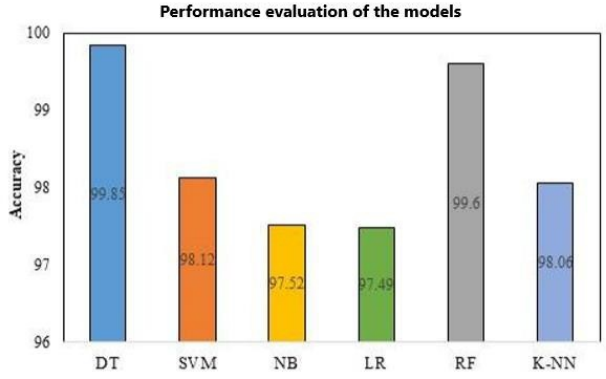

Figure 4. Classifying Best Algorithm of Machine Learning for COVID-19 infected patients' data set by Measuring the Accuracy of the Models

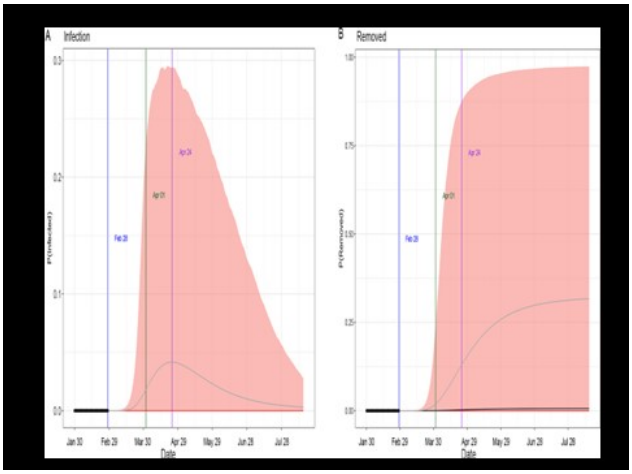

Figure 5. COVID 19 Prediction trend Under No Precautionary Measures in China according to the SIR model

Machine learning algorithm is best to forecast the patient's recovery.

1) Performance Evaluation of Model For COVID-19

Figure 4 illustrated the occurrence of each element of the dataset. The Machine Learning algorithm is calculated using the evaluation technique to determine the accuracy of the classifier [18]. This method determines the effectiveness, quality, and efficiency of the model with help of a machine learning algorithm. Performance evaluation techniques for machine learning include sensitivity, specificity, precision, and accuracy. However, we use accuracy to evaluate the overhead model. From the Figure 4, the model developed with Decision A tree has the highest accuracy rate of 99.85 per cent, second by a Random Forecast 99 percent, a Naïve Bay with 97.52 percent accuracy, a 97.49 percent accuracy logistic recovery, a 98.85 percent precision vector support, then a K-NN with98.06

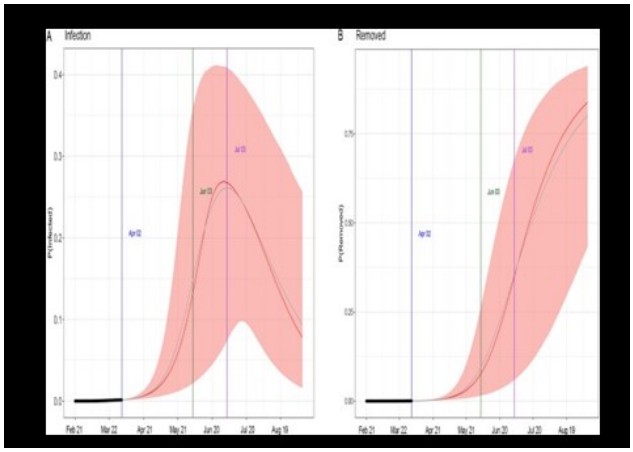

Figure 6. COVID 19 Prediction trend Under No Precautionary Measures in Italy according to the SIR model

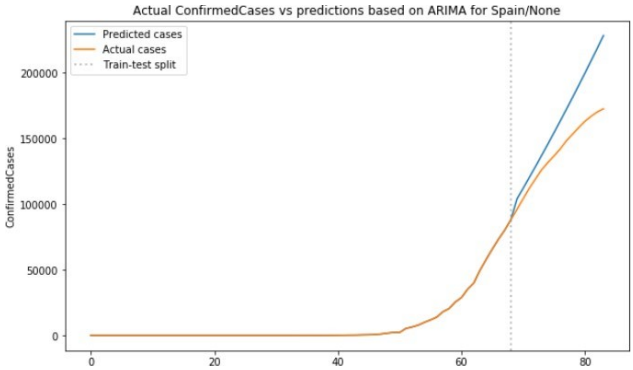

Figure 7. Prediction of Actual and Confirmed cases of COVID 19 using ARIMA model in Spain

2) COVID-19's prediction Using SIR Model for Italy and China

From Figure 5, we can see the tendency of COVID-19 prediction under no preventative measures in China. We can see the black dots on the left side of the blue line which determines the proportion of those people who are infected upon the last date of availability or earlier that. The line in blue color in the above figure demonstrates the time-independent variable t0. Vary from the purple and green lines which shows the first and the second points of turning respectively. The area within the 95 percent credible interval consists of cyan and salmon color which is very predictable for the proportions of infectious diseases before and after. For portion (A), the turning times for disease transmission From Figure 6, after June 3 virus rate starts increasing at a high rate and touching the maximum limit. Also, after July 3 virus rate started decreasing.

3) ARIMA Model for Predicting COVID-19 in Spain, Italy, and China

The overhead Figure 7 clarifies the situation of the COVID-19 Actual and Predicted cases in Spain. On the X-Axis we can see the time duration of the COVID-19 starting

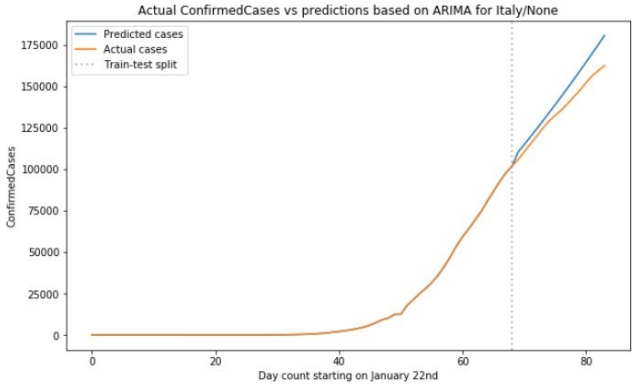

Figure 8. Prediction of Actual and Confirmed cases of COVID 19 using ARIMA model in Italy

from January 22, 2020, to onward. And on the other hand, on Y-Axis number of confirmed cases are given. The Blue Line outlines the projected instances in Spain, whereas the Orange Line shows the genuine COVID-19 cases in Spain. The above data confirms that the time series is not stagnant and that the series data plot above shows an increasing trend that suggests that COVID-19 instances have increased significantly. The actual and predicted specific COVID-19 cases for the fair comparison are a time series from 22 January 2020 upwards. The resemblance between the forecasted data and actual data is clear from the plots. The above comparison reveals the precision of the ARIMA model for forecasting purposes.

The Figure 8 predicts the actual and confirmed cases of COVID-19 using the ARIMA model on the data set of Italy. On the X-Axis we can see the time of the COVID-19 starting from January 22, 2020, to onward. And on the other hand, on the Y-Axis number of confirmed cases are shown. The Blue labeled line illuminates the predicted cases in Italy and the Orange line is illustrating the actual number of COVID-19 cases in Italy. Also, can see the time series is not static. These charts show the resemblance of estimated data to actual data. This comparison reveals the model's accuracy, sensitivity and accuracy.

5. CONCLUSION AND FUTURE WORK

Machine learning includes Decision Tree, Support Vector Machine, Naïve Bayes, Random and K-NN logistic regression. These are all directly applied with Python on the COVID-19 dataset. However, with the best accuracy of 99.85 percent the model performance employing the Decision Tree Algorithm appears to be the finest in most other models. Decision Tree algorithm is accomplished to forecast the probability of recovered infectious patients from COVID-19 with an overall accuracy of 99.85 percent when compared with Random Forecast, Support Vector Machine, K-NN, Naïve Bayes, and Logistic Regression with the overall accuracy of 99.60 percent, 98.85 percent, 98.06 percent, 97.52 percent, and 97.49 percent respectively. COVID-19 is putting unprecedented mental stress over the public health sector of all countries. Countries confronting the foremost danger are the USA, Italy, and Spain. The efficiency of partial and complete lockdown is still unknown. Under this situation, we have implemented the SIR model. Among the advantages of the implementing SIR model, it is worth noting that despite the simplicity of this hypothesis, the adjustments obtained are accurate and the predictions made do not differ much from those other more complex models. Here we have compared the situation of China with Italy as they have kind of same population characteristics to make our predictions using the SIR model. The emergence of COVID-19 seemed relatively rapid in the Hunan province. Since the comparison between population size and geographic proximity in the area of Hubei, public health operations in Hunan can be useful and beneficial in preventing Italy from spreading COVID-19 more effectively. In this Study, we used the SIR model using Python Programming language to calculate the influence of intervention measures of the COVID-19 epidemic in Italy. The advantages of the System model are that it combines time variable insulation measures with the SIR model to modify the time-variable transmission rate in the population [18]. COVID-19 spread outside of Hubei China was less severe but they did not perform analysis in each province. The first and second points of the survey are February 04, 2020, and February 09, 2020, which are the same as are outside of Hubei, China. The confirmed number of infected cases is 1,018 which are included in the forecasted infectious cases (13, 840–8) to the forecasted endpoint March 14, 2020, which are included in our research. This also demonstrates the consistency, precision, sensitivity, and accuracy of the SIR model. When we combine the above datasets through SIR Model, the above findings and outcomes demonstrate the SIR model is most appropriate for calculating the epidemic trend of COVID- 19. The previous studies show that strict government policies resulted in a slower rise in the infected population [18]. In our research, we study those rigid policies radically reduce the cases of COVID-19. According to our SIR model, Italy still maintains its precautionary measures, isolation, and quarantines like China. Due to the above parameters, both countries reduce their cases. Our study compares both countries' rigid governmental controlled policies. Travel restrictions, tracking of the individuals, and entry, exit measures at all points could decrease the spread of the infection. World Health Organization introduces some precautionary measures such as personal hygiene, environmental cleanliness, travel restrictions, etc. Strict Public health Precautionary Measures reduce cases all over the world. It should be possible that we don't have the exact number of cases because of the unverified and unheeded cases. The incubation period is not included in this study. Because some scholars said that if we ignore the incubation period, the SIR model is possibly being overestimated. The accuracy of the SIR model could be increased by adding an incubation period for better predictions of the SIR model.

6. ACKNOWLEDGMENT

The authors are highly thankful to Chairperson, Department of Information Systems for their valuable guidance and support.

References

[1] K. Xiao, J. Zhai, Y. Feng, N. Zhou, X. Zhang, J.-J. Zou, N. Li, Y. Guo, X. Li, X. Shen *et al.*, "Isolation and characterization of 2019-ncov-like coronavirus from malayan pangolins," *BioRxiv*, 2020.

[2] W. Ji, W. Wang, X. Zhao, J. Zai, and X. Li, "Homologous recombination within the spike glycoprotein of the newly identified coronavirus may boost cross-species transmission from snake to human," *Journal of medical virology*, 2020.

[3] W. Qiu, C. Chu, A. Mao, and J. Wu, "The impacts on health, society, and economy of sars and h7n9 outbreaks in china: a case comparison study," *Journal of environmental and public health*, vol. 2018, 2018.

[4] H. Nisar, B. Wajid, S. Shahid, F. Anwar, I. Wajid, A. Khatoon, M. U. Sattar, and S. Sadaf, "Whole-genome sequencing as a first-tier diagnostic framework for rare genetic diseases," *Experimental Biology and Medicine*, p. 15353702211040046, 2021.

[5] A. Ghaffar, S. Alanazi, M. Alruwaili, M. U. Sattar, W. Ali, M. Humayun, S. Y. Siddiqui, F. Ahmad, and M. A. Khan, "Multi-stage intelligent smart lockdown using sir model to control covid 19," *Intelligent Automation and Soft Computing*, pp. 429–445, 2021.

[6] Z. Zahid, M. U. Sattar, H. W. Khan, A. Zahid, and M. F. Riaz, "A smart analysis and visualization of the power forecasting in pakistan," *International Journal of Computing and Digital Systems*, vol. 10, 2021.

[7] M. Narmeen, M. U. Sattar, M. Fatima, H. W. Khan, M.-u.-D. Azad, and F. Ghani, "Impact of weather on covid-19 in metropolitan cities of pakistan: A data-driven approach," *International Journal of Computing and Digital System*, 2021.

[8] A. S. S. Rao and J. A. Vazquez, "Identification of covid-19 can be quicker through artificial intelligence framework using a mobile phone–based survey when cities and towns are under quarantine," *Infection Control & Hospital Epidemiology*, vol. 41, no. 7, pp. 826–830, 2020.

[9] L. Yan, H.-T. Zhang, Y. Xiao, M. Wang, Y. Guo, C. Sun, X. Tang, L. Jing, S. Li, M. Zhang *et al.*, "Prediction of criticality in patients with severe covid-19 infection using three clinical features: a machine learning-based prognostic model with clinical data in wuhan," *MedRxiv*, 2020.

[10] Y. Li, M. Liang, X. Yin, X. Liu, M. Hao, Z. Hu, Y. Wang, and L. Jin, "Covid-19 epidemic outside china: 34 founders and exponential growth," *Journal of Investigative Medicine*, vol. 69, no. 1, pp. 52–55, 2021.

[11] J. F. Drexler, V. M. Corman, and C. Drosten, "Ecology, evolution and classification of bat coronaviruses in the aftermath of sars," *Antiviral research*, vol. 101, pp. 45–56, 2014.

[12] B. Hu, X. Ge, L.-F. Wang, and Z. Shi, "Bat origin of human coronaviruses," *Virology journal*, vol. 12, no. 1, pp. 1–10, 2015.

[13] B. Ahmad, "Intelligent digital twin to make robot learn the assembly process through deep learning," *LGURJCSIT*, vol. 5, no. 3, pp. 65–72, 2021.

[14] D. Mining, "Concepts and techniques," *Jiawei Han and Micheline Kamber*, vol. 2, 2001.

[15] K. Deepika and S. Seema, "Predictive analytics to prevent and control chronic diseases," in *2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT)*. IEEE, 2016, pp. 381–386.

[16] H. Asri, H. Mousannif, H. Al Moatassime, and T. Noel, "Using machine learning algorithms for breast cancer risk prediction and diagnosis," *Procedia Computer Science*, vol. 83, pp. 1064–1069, 2016.

[17] M. M. Islam, H. Iqbal, M. R. Haque, and M. K. Hasan, "Prediction of breast cancer using support vector machine and k-nearest neighbors," in *2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC)*. IEEE, 2017, pp. 226–229.

[18] L. Muhammad, A. A. Haruna, I. A. Mohammed, M. Abubakar, B. G. Badamasi, and J. M. Amshi, "Performance evaluation of classification data mining algorithms on coronary artery disease dataset," in *2019 9th International Conference on Computer and Knowledge Engineering (ICCKE)*. IEEE, 2019, pp. 1–5.

Muhammad Imran is a graduate student in Information System Department at University of Management and Technology.

Usman Sattar has more than fifteen years of teaching, training, and consultancy experience in the fields of Information System especially Immersive Technologies, Data Science, and Enterprise Systems System Security. He is PhD (Informatics) – Scholar at Malaysia University of Science and Technology, Malaysia.

Hamza Wazir Khan has seven years of diversified experience in field of Information System and Marketing. Currently, he is working as a Lecturer in Information System Department at University of Management and Technology. He has given his services in form of teaching, training, and consultancy in fields of Data Science, Digital Marketing, Enterprise Resource Planning, and E Business.

Abdul Ghaffar a PhD candidate at University of Management and Technology, Lahore, alumni from University of the Punjab, has a diverse background of Management and soft skills. He is associated with the field of training, development and research since the last 20 years. Mr. Ghaffar's major research interests are HRM, Organizational Psychology / behavior (especially job burnout, job stress, psychological capital, emotional labor etc.

Dr. Hammad Mushtaq holds PhD from WUT, China and a MS from NUS, Singapore. His research work includes Digital Marketing, online consumer behavior, Enterprise Systems adoption and Business Intelligence.