International Journal on Emerging Technologies 12(1): 335-341(2021)

ISSN No. (Print): 0975-8364
ISSN No. (Online): 2249-3255

Analyzing the Epidemiological Outbreak of COVID-19 using Descriptive and Predictive Analytics

Hakeem Ur Rehman[1], Muhammad Shafiq[1], Fizza Ghulam Nabi[1], Sana Hassan[1], Sajjad Ahmad Baig[2] and Umair Manzoor[2]*

[1]*University of the Punjab, Lahore, Pakistan.*
[2]*National Textile University, Faisalabad, Pakistan.*

(Corresponding author: Sajjad Ahmad Baig)
(Received 02 January 2021, Revised 03 February 2021, Accepted 11 March 2021)
(Published by Research Trend, Website: www.researchtrend.net)

ABSTRACT: The basic purpose of this study is to give a detailed insight view on COVID-19 epidemic. Recent outbreak of COVID-19 has spread to more than 50 countries which was first reported in Wuhan, China. COVID-19 has been declared as Public Health Emergency of International Concern (PHEIC) on 30 January 2020 by World Health Organization. Fast spreading is mostly involved naturally when a rising infectious disease outbreaks which endangers the health of many people requiring immediate actions to prevent the disease.

The theoretical and empirical results of this study have important contribution in literature of this field. Data is collected from an online website named "Kaggle.com". We did descriptive analysis and predictive analysis through Python to find a detailed insight of the current situation and to predict cases trends in the coming days.

This dataset can be useful to monitor the emerging trends in 2019-nCoV. Dissemination of the detailed data from this paper to the scientific community can be very helpful when there is a little else available. We investigated early indications that the response is being strengthened in China and worldwide based on a decrease in the case of detection time and rapid management worldwide. This is an early data analysis and visualization approach of a situation that is rapidly evolving.

Keywords: COVID-19, Descriptive Analytics, Predictive Analytics

I. INTRODUCTION

Recent outbreak of COVID-19 has spread to more than 50 countries [14] which was first reported in Wuhan, China. COVID-19 has been declared as Public Health Emergency of International Concern (PHEIC) on 30 January 2020 by World Health Organization (WHO) [23, 25]. Fast spreading is mostly involved [15] naturally when a rising infectious disease outbreaks which endangers the health of many people requiring immediate actions to prevent the disease. 2019-nCoV belongs to the subfamily of Orthocoronavirinae, different from severe acute respiratory syndrome coronavirus and Middle East respiratory syndrome-coronavirus [29]. On 12th December 2019, the first case of an unexplained new pneumonia origin was detected [28, 29]. Later it was found as a non-SARS nCoV by the Chinese Center for Disease Control and Prevention (CDC). This virus is known to cause the common cold and diarrheal diseases in humans, belonging from Coronaviridae family which consists plus stranded RNA viruses, single and large isolated multiple species [6,7,18, 24]. According to the state council information office in Beijing, China's capital, 26 January 2020, in china, a total of 1975 cases of pneumonia have been confirmed [10, 15,16, 20, 21]. Due to virus the government of china has been extreme careful, yet the virus has tended to spread out of China [1, 8]. Since the outbreak Peru reported 98,377 cases, Ecuador reported 33,956 cases, Colombia reported 85,467 cases, Chile reported 23,204 cases, Brazil reported 6,43,430 cases, USA reported 18,05,739, Mexico reported 81,183 cases, Canada reported 27,518 cases, Russia reported 2,15,508 cases, Italy reported 12,919 cases, Germany reported 6,122 cases, France reported 63,751 cases, Belgium reported 35,771 cases, Turkey reported 12,871 cases, Iran reported 23,692 cases, India reported 3,20,339 cases while Pakistan reported 77,573 cases [12].

Table 1.

Country/Region	Active Cases
India	320339
Iran	23692
Pakistan	77573
Turkey	12871
Belgium	35771
France	63751
Germany	6122
Italy	12919
Russia	215508
Canada	27518
Mexico	81183
USA	1805739
Brazil	643430
Chile	23204
Colombia	85467
Ecuador	33956
Peru	98377

Source: (Kp, 2020)

In November, many patients reported to have visited a local fish and wild animal market in Wuhan, so the reason behind the origin of epidemic was the animal to human transmission. Apart from this study, animal to the human and interhuman transmission of the virus has been studied recently [11, 16]. To manage and control this epidemic, we need to have a better understanding of its pandemic nature because situation is getting serious day by day [13, 14,19]. Our research team believes that these epidemic data should be openly available and easily accessible for all health professionals and data scientists. This dataset would serve as a base for people to gather more data about epidemics. This study aims to provide data relevant to COVID-19 epidemic so it would play its part in cause analysis and management strategies. Therefore, kaggle was born as the online platform that provides latest and reliable news development, as well as statistics on COVID-19. This paper aims to predict and forecast COVID19 cases, deaths, and recoveries through predictive modelling.

II. REVIEW

In 2019 the COVID-19 coronavirus disease, which firstly originates in Wuhan, China, has infecting 213 territories and countries throughout the world (WHO) [2]. In Wuhan the Cases of COVID-19 were treated as pneumonia appeared with unknown causes, in December 2019 and spread in all 213 countries across the Globe with high rate. Although most of the infected patients show minor symptoms such as fever, symptoms of track upper respiratory system, difficulty in breathe [9], as well the asymptomatic [3], and the severely infected patients lead to pneumonia, failure of multiple body organs and death [4, 26]. Almost 808,995 peoples have been dead due to COVID-19 globally [9], and it is still expected to raise the number of deaths with condition of ongoing epidemic.

The significant health threat posed by COVID-19 to humans attracts high attention from public health researchers around the world, and more than total 200 publications in different academic journal during last 2 months. Also, the total of 150 papers on Covid-19 are almost “epidemiological and preprint modeling”, and most of the scholars want to explore the epidemiological disease parameters at various different countries vide locations to disseminate the most critical info among between policymakers as well as the modelers to bring timely implement the response control. This thins is especially important as the latest outbreak of COVID-19 involve a modern pathogen of “SARS-CoV-2”, on which little information’s available of its clinical profile and infectivity. The recent estimation from this investigation are very wide, significantly the differences in Analytical tools and assumptions. The latest estimation of this study is very broad, somewhat because of the variation I the method of analysis and setting assumptions. This type of variations must be reflected during the evaluation of effectiveness of the intervention of public health globally implemented. As the outbreak of COVID-19 is at a conclusive point as suggested by the “Director-General of the WHO” on February 27, 2020; “it is imperative to synthesize all existing evidence available to date and summarize the key findings to identify research gaps and to assist policymakers in evidence-based decision making for better pandemic preparedness”. The COVID-19 Outbreak is going to be fast in some countries as the already infected patients cause to infect the 2 to 3 humans at average and the ratio will be double within three to seven days. Somehow the duration of incubation ranged between 3-6 days depending on the different published sources and preprints modules. It suggest that if we use median than the number of days for incubation likely to be 5 days at an average rate, it indicate that COVID-19 outbreak and spreading rate is similar to that of other corona viruses, like “SARS-CoV (4.4 days)” [17] and “MERS-CoV (5.5 to 6.7 days)” [5, 25].

III. METHODOLOGY

The basic purpose of this study is to give a detailed insight view on COVID-19 epidemic. Another purpose is to predict data and its impact in future. Both the theoretical and empirical results of this study have important contribution in literature of this field.

Data. Data is collected from an online website named “Kaggle.com”.

IV. DATA ANALYSIS

We did descriptive analysis and predictive analysis through Python to find a detailed insight of the current situation and to predict cases trends in the coming days. Descriptive analysis helped us to understand the current situation regarding COVID-19 reported cases, COVID-19 tests conduction and deaths occurred from this pandemic. Predictive analysis helped us to predict the future state of the pandemic.

A. Data visualization and Descriptive Analytics

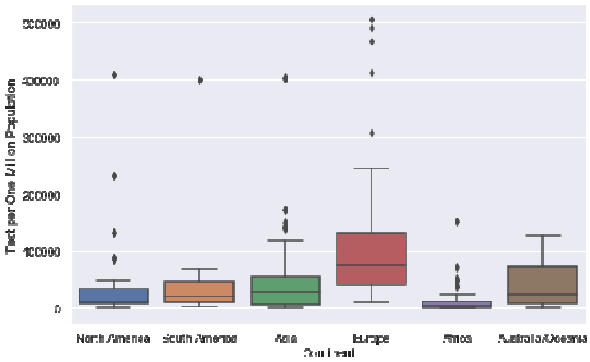

Fig. 1. COVID19 tests per One-Million Population conducted in each Continent.

Descriptive analysis was conducted to figure out the number of tests conducted, the total cases and number of death cases happening all over the world in each continent. Figure. 1 depicts number of tests conducted in each continent per 1 million population. In North America, South America, Asia, Africa and Australia, less than 1 lac tests were conducted per one million population but in the European region around 1 lac 25 thousand tests were conducted per one million population. Due to rapid growth of COVID-19 cases, in North America it is expected to conduct 90,000 tests initially later around 1,30,000 than around 2,30,000 and in the worst scenario above 4,00,000 tests can be conducted per one million population, South America and Asia expects around 4,00,000 tests conduction, Europe initially expects conduction of 3,00,000 tests which can later increase to above 5,00,000 with the changing scenario, Africa expects 80,000 tests initially

which can increase up to 1,70,000 tests and Australia expects around 1,20,000 – 1,30,000 tests conduction.

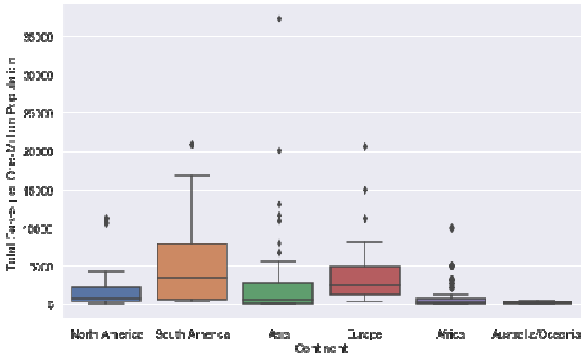

Fig. 2. COVID19 cases per One-Million Population in each Continent.

Fig. 2. represents the number of corona cases per one million population in each continent. It shows that less than 5000 cases were reported in North America, Asia, Africa, and Australia. In North America the case are expected to grow between 11,000 – 12,0000 from 5,000, Asia has possibility to grow 14,000 cases rapidly later increasing the expected cases to 20,000 and in the worst scenario the amount could reach about 35,000 cases per one million population.

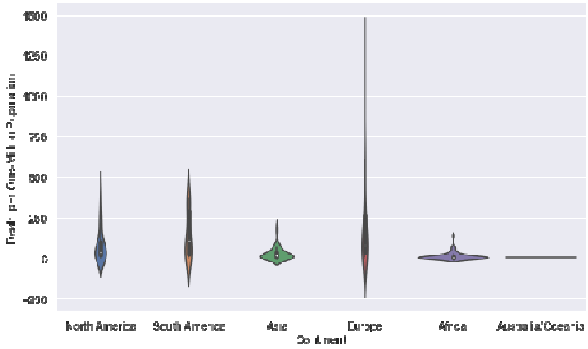

Fig. 3. Deaths per One-Million Population in each Continent.

Africa reported around 1,000 cases which are expected to 5,000 and in the worst scenario the amount of cases could reach to 10,000, Australia/Oceania reported less than thousand cases which are under control. If we talk about South America, it reported around 8,000 cases per one million population which are expected to grow nearly 21,000 cases. Europe reported 5,000 cases per one million population which are expected to grow 10,000 initially, later 15,000 and in the worst scenario around 21,000 cases per one million population. Fig. 3. represents the number of death cases per one million population in each continent. It shows that less than 200deaths were reported in North America which are expected to grow above 400 death cases, less than 100 deaths were reported in Asia which are expected to grow with 200 death cases, less than 100 deaths are reported in Africa which are expected to grow around 200, less than 50 death cases were reported in Australia and the situation is under control, In South America the death cases are reported around 400, Europe reported above 200 death cases which are expected to increase by 600 cases and later above 1200 cases in the worst scenario.

Table 2 shows a detailed view of active cases in each continent, also in detail classifying the cases into 5 major categories including: 1) Serious cases 2) Total deaths 3) Total recovered 4) Total cases and 5) Total tests conducted till now. Africa has reported 629142 cases, Australia/Oceania reported 12154 cases, Europe reported 2599901 cases, North America reported 4153009 cases, South America reported 2993241 cases while Asia reported 2992577 cases until now.

Table 3 gave an overall continent view about the COVID-19 cases. In table 2, countries who reported more than 5000 active COVID-19 cases and deaths has been mentioned. Brazil, Chile, Colombia, Ecuador, and Peru are some of the regions from South America who have more than 5000 active cases and deaths. Similarly, (Canada, Mexico, USA) from North America, (Belgium, France, Germany, Italy, Russia) from Europe, (India, Iran, Turkey and Pakistan) from Asia are the regions who reported more than 5000 active cases and death cases.

Table 2: Total active, serious/critical, deaths, recovered, COVID19 cases and tests in each continent.

Continent	Total Active Cases	Serious / Critical	Total Deaths	Total Recovered	Total Cases	Total Tests
Africa	297623	982	13803	317716	629142	6054121
Asia	842791	19782	68120	2081666	2992577	44907122
Australia/Oceania	2488	28	133	9533	12154	3672053
Europe	440235	5319	197291	1325213	2599901	73480987
North America	2024350	19737	188994	1939665	4153009	49172998
South America	989205	13640	108399	1895637	2993241	11303001

Table 3: Continent wise list of countries having more than 5000 deaths and active cases.

Continent	Country/Region	Active Cases	Serious/Critical	Total Deaths
Asia	India	320339	8944	24327
	Iran	23692	3389	13211
	Pakistan	77573	2078	5386
	Turkey	12871	1204	5402
Europe	Belgium	35771	23	9787
	France	63751	492	30029
	Germany	6122	264	9144
	Italy	12919	60	34984
	Russia	215508	2300	11614
North America	Canada	27518	2170	8798
	Mexico	81183	378	36327
	USA	1805739	16337	139143
South America	Brazil	643430	8318	74262
	Chile	23204	1915	7069
	Colombia	85467	875	5625
	Ecuador	33956	313	5130
	Peru	98377	1325	12229

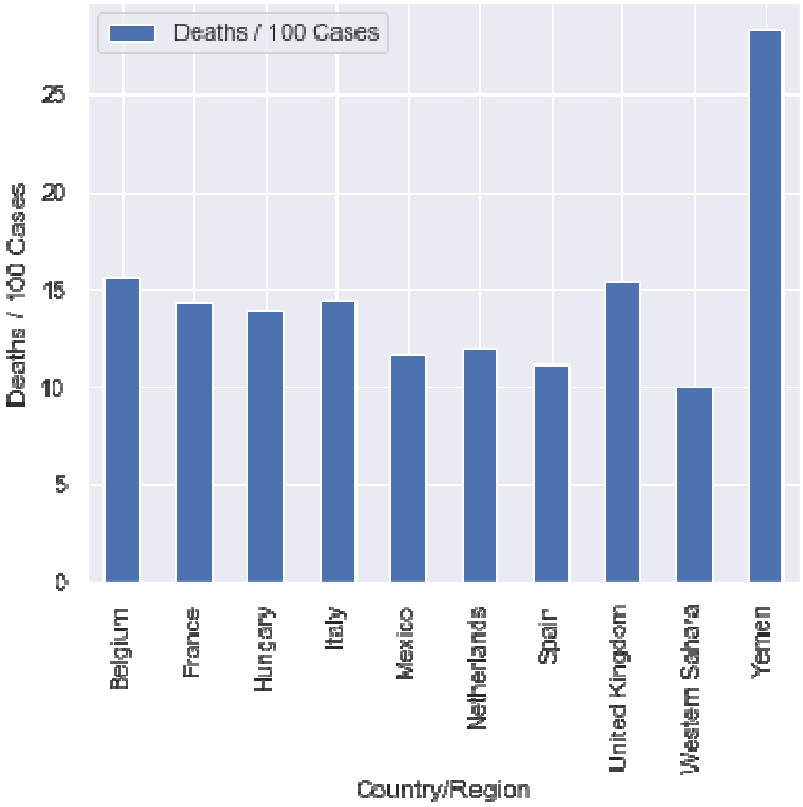

Fig. 4. Top ten highest death rate (deaths per 100 cases) countries due to COVID19.

Countries having highest death rate due to COVID-19 has been shown in table 4. Yemen, Belgium, UK, Italy, France, Hungary, Netherlands, Mexico, Spain and Western Sahara are the top 10 countries suffering from highest death rate respectively.

Top 10 countries who have highest recovery rate have shown in table 5. Brunei, Dominica, Greenland, Hole see, Iceland, Malaysia, Malta, New Zealand, Qatar, Taiwan are the top 10 countries with highest recovery rate ranging from 95 to 100 recoveries out of 100 cases.

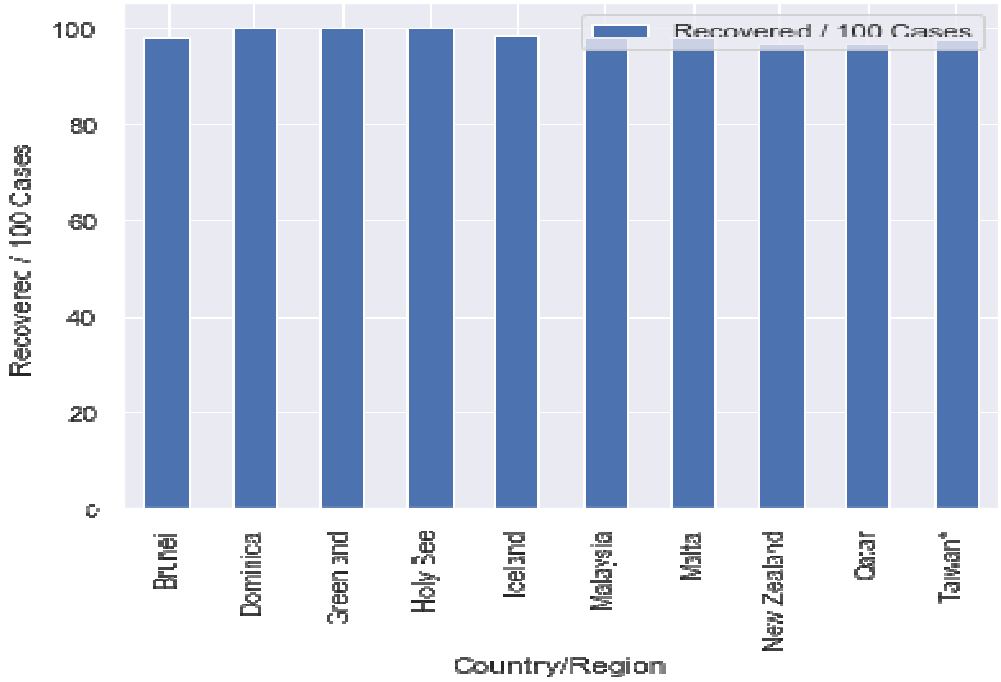

Fig. 5. Top ten highest recovery rate (recovered per 100 cases) countries due to COVID19.

Predictive Analytics:
Prediction:

Table 4: Holt's Exponential Smoothing Method with linear trend.

	Name	Optimum Parameter Value
smoothing_level	alpha	0.93
smoothing_slope	beta	0.21

Holt in 1957extended simple exponential smoothing to allow the forecasting of data with a trend. In this paper Holt's Exponential Smoothing Method was used to predict COVID-19 cases trends. In table 4, The value of apha (α = 0.93) shows immediate rapid growth in death cases due to COVID-19 which means when the pandemic outbreak happened, the death cases increased significantly. The value pf Beta (ß=.21) shows that there is a less probability of the death cases due to COVID-19 in the future.

Table 5: Death Rate due to COVID19 in the World (deaths per 100 cases) (July, 2020).

15th	16th	17th	18th	19th	20th	21st	22nd	23rd	24th
4.30	4.27	4.23	4.19	4.16	4.12	4.08	4.05	4.01	3.97

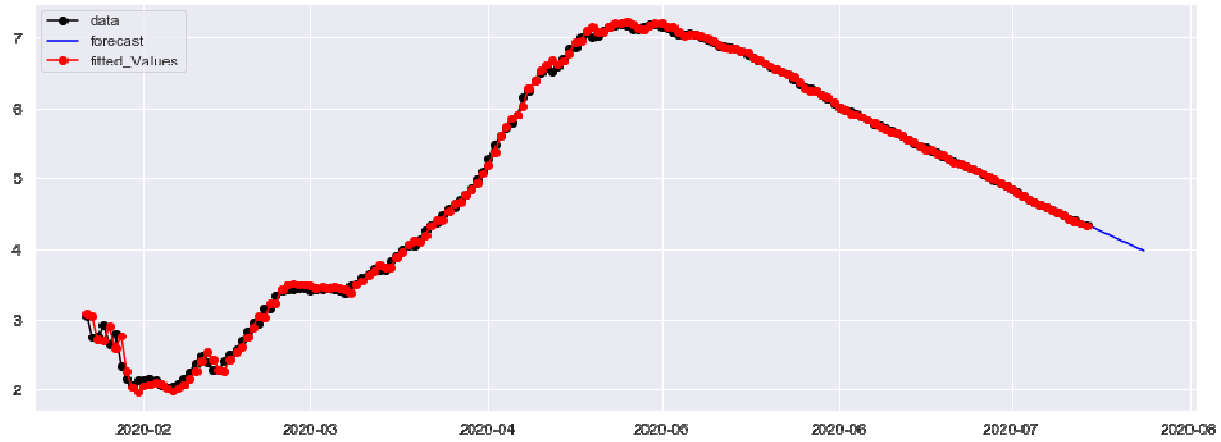

Fig. 6. Death per 100 cases in the world due to COVID19.

Recent trend of deaths cases from COVID-19 has been studied and shown in figure 6. The data shows that average death worldwide from COVID-19 would reduce to 4 persons per 100 cases.

Table 6: Holt's Exponential Smoothing Method with linear trend.

	Name	Optimum Parameter Value
smoothing_level	alpha	0.74
smoothing_slope	beta	0.74

The value of apha (α = 0.74) shows recovery ratefrom COVID-19 which means there is 74% probability of recovery from COVID-19. The value pf Beta (ß=.74) shows that there is a steady probability of recovery from COVID-19 in the future.

Table 7: COVID19 patients Recovery Rate in the World (recover per 100 cases) (July, 2020).

15th	16th	17th	18th	19th	20th	21st	22nd	23rd	24th
54.98	55.18	55.37	55.56	55.75	55.94	56.13	56.33	56.52	56.71

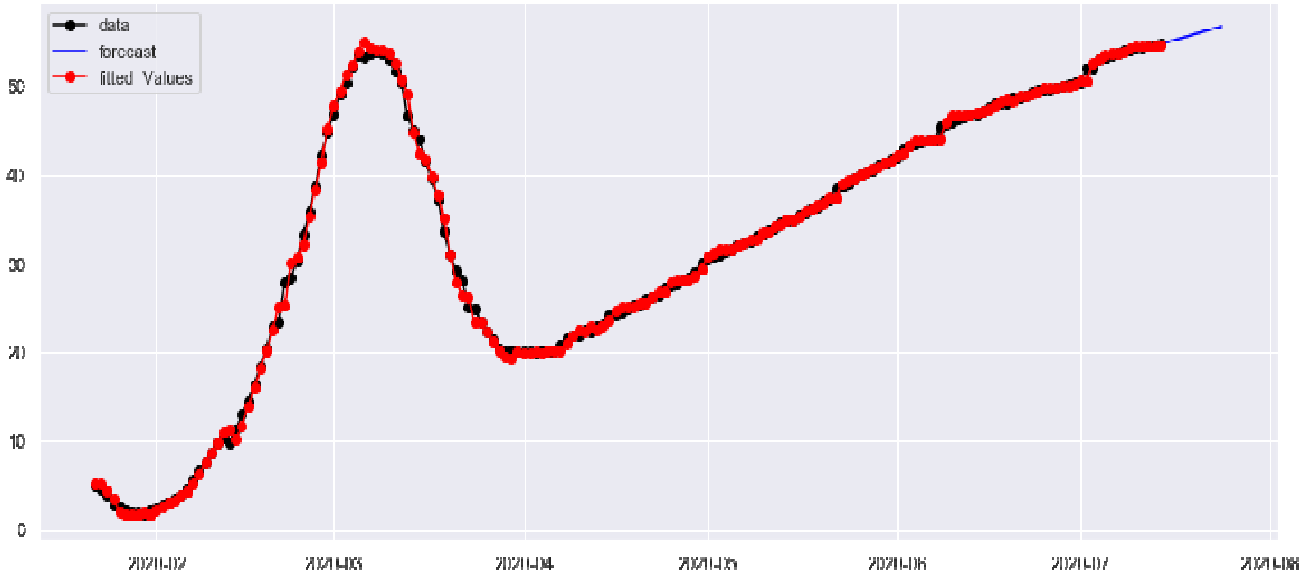

Fig. 7. Recovered per 100 cases in the world due to COVID19.

Recent trend of recovered cases from COVID-19 has been studied and shown in fig. 7. The data shows that average recovery worldwide from COVID-19 is somewhere between 54% to 56% which means 54% to 56% of the cases will recover in future.

CONCLUSION

In, conclusion, we have used the dataset from 15 July 2020 to 24 July 2020 for our 2019 coronavirus experiment. This dataset can be useful to monitor the emerging trends in 2019-nCoV. Dissemination of the detailed data from this paper to the scientific community can be very helpful when there is a little else available. We investigated early indications that the response is being strengthened in China and worldwide based on a decrease in the case of detection time and rapid management worldwide. We observed an interesting fact, which helps us to understand that it needs to be managed until the vaccine is made. This is an early data analysis and visualization approach of a situation that is rapidly evolving. Hopefully, in future, we will continue to monitor this outbreak's data that we have used in this study and from other official sources.

REFERENCES

[1]. Azam, T., Mohsin, M., Naseem, S., Nilofar, M., Zia-Ur-Rehman, M., Nelofer, S., et al. (2020). Economic growth vulnerability amid the COVID-19 epidemic: a systematic review of different sectors of Pakistan. Rev. Argentina Clín. Psicol. 29, 705–713. doi: 10.24205/03276716.2020.875

[2]. BNO Tracking Coronavirus: Map, Data and Timeline. Available online: https://bnonews.com/index.php/ 2020/02/the-latest-coronavirus-cases/ (accessed on 18 March 2020).

[3]. Bai, Y.; Yao, L.; Wei, T.; Tian, F.; Jin, D.Y.; Chen, L.; Wang, M. Presumed Asymptomatic Carrier Transmission of COVID-19. JAMA 2020. [CrossRef] [PubMed]

[4]. Cauchemez, S.; Fraser, C.; Van Kerkhove, M.D.; Donnelly, C.A.; Riley, S.; Rambaut, A.; Enouf, V.; van der Werf, S.; Ferguson, N.M. Middle East respiratory syndrome coronavirus: Quantification of the extent of

the epidemic, surveillance biases, and transmissibility. Lancet Infect. Dis. 2014, 14, 50–56.
[5]. Cowling, B.J.; Park, M.; Fang, V.J.; Wu, P.; Leung, G.M.; Wu, J.T. Preliminary epidemiological assessment of MERS-CoV outbreak in South Korea, May to June 2015. Euro Surveill. 2015, 20, 7–13.
[6]. Chen, Y., Liu, Q., & Guo, D. (2020). Emerging coronaviruses: genome structure, replication, and pathogenesis. Journal of Medical Virology, 92(4), 418–423.
[7]. Drosten, C., Günther, S., Preiser, W., Van Der Werf, S., Brodt, H.-R., Becker, S., Rabenau, H., Panning, M., Kolesnikova, L., & Fouchier, R. A. M. (2003). Identification of a novel coronavirus in patients with severe acute respiratory syndrome. New England Journal of Medicine, 348(20), 1967–1976.
[8]. Gates, B. (2020). New engla nd journal. 1677–1679.
[9]. Guan, W.J.; Ni, Z.Y.; Hu, Y.; Liang, W.H.; Ou, C.Q.; He, J.X.; Liu, L.; Shan, H.; Lei, C.L.; Hui, D.S.C.; et al. Clinical Characteristics of Coronavirus Disease 2019 in China. N. Engl. J. Med. 2020.
[10]. Hui, D. S., Azhar, E. I., Madani, T. A., Ntoumi, F., Kock, R., Dar, O., Ippolito, G., Mchugh, T. D., Memish, Z. A., & Drosten, C. (2020). The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health—The latest 2019 novel coronavirus outbreak in Wuhan, China. International Journal of Infectious Diseases, 91, 264–266.
[11]. Ji, W., Wang, W., Zhao, X., Zai, J., & Li, X. (2020). Cross-species transmission of the newly identified coronavirus 2019-nCoV. Journal of Medical Virology, 92(4), 433–440.
[12]. Kp, D. (2020). COVID-19 Dataset.
[13]. Lauren G. (2020). Coronavirus COVID-19 Global Cases.
[14]. Lipsitch, M., Swerdlow, D. L., & Finelli, L. (2020). Defining the epidemiology of Covid-19 - Studies needed. New England Journal of Medicine, 382(13), 1194–1196. https://doi.org/10.1056/NEJMp2002125
[15]. Liu, Y., Gayle, A. A., Wilder-Smith, A., & Rocklöv, J. (2020). The reproductive number of COVID-19 is higher compared to SARS coronavirus. Journal of Travel Medicine, 27(2), 1–4. https://doi.org/10.1093/jtm/taaa021
[16]. Lu, H., Stratton, C. W., & Tang, Y. (2020). Outbreak of pneumonia of unknown etiology in Wuhan, China: The mystery and the miracle. Journal of Medical Virology, 92(4), 401–402.
[17]. Lau, E.H.; Hsiung, C.A.; Cowling, B.J.; Chen, C.H.; Ho, L.M.; Tsang, T.; Chang, C.W.; Donnelly, C.A.; Leung, G.M. A comparative epidemiologic analysis of SARS in Hong Kong, Beijing and Taiwan. BMC Infect. Dis. 2010, 10, 50.
[18]. Mehta, P., McAuley, D. F., Brown, M., Sanchez, E., Tattersall, R. S., & Manson, J. J. (2020). COVID-19: consider cytokine storm syndromes and immunosuppression. The Lancet, 395(10229), 1033–1034. https://doi.org/10.1016/S0140-6736(20)30628-0
[19]. Mohsin, M., Naiwen, L., Zia-UR-Rehman, M., Naseem, S., & Baig, S. A. (2020b). The volatility of bank stock prices and macroeconomic fundamentals in the Pakistani context: an application of GARCH and EGARCH models. Oeconomia Copernicana, 11(4), 609-636.
[20]. Mohsin, M., Naseem, S., Zia-ur-Rehman, M., Baig, S. A., & Salamat, S. (2020a). The crypto-trade volume, GDP, energy use, and environmental degradation sustainability: An analysis of the top 20 crypto-trader countries. International Journal of Finance & Economics.
[21]. Naseem, S., Mohsin, M., Hui, W., Liyan, G., & Penglai, K. (2021). The Investor Psychology and Stock Market Behavior During the Initial Era of COVID-19: A Study of China, Japan, and the United States. Frontiers in Psychology, 12, 16.
[22]. Sahin AR, Erdogan A, Agaoglu PM, Dineri Y, Cakirci AY, Senel ME, Tasdogan AM. 2019 novel coronavirus (COVID-19) outbreak: a review of the current literature. EJMO 2020;4(1):1–7.
[23]. Peng, P. W. H., Ho, P. L., & Hota, S. S. (2020). Outbreak of a new coronavirus: what anaesthetists should know. British Journal of Anaesthesia, 124(5), 497–501. https://doi.org/10.1016/j.bja.2020.02.008
[24]. Report, M. W. (2020). Severe Outcomes Among Patients with Coronavirus Disease 2019 (COVID-19) — United States, February 12–March 16, 2020. 2020;69:343-346. DOI: http://dx.doi.org/10.15585/mmwr.mm6912e2. MMWR Morb Mortal Wkly Rep, *69*(12), 343–346. https://doi.org/10.15585/mmwr.mm6912e2
[25]. Report, S. (2020). Coronavirus disease 2019 (COVID-19) Situation Report-72 HIGHLIGHTS. 2019(April).
[26]. Salamat, S., Lixia, N., Naseem, S., Mohsin, M., Zia-ur-Rehman, M., & Baig, S. A. (2020). Modeling cryptocurrencies volatility using GARCH models: a comparison based on Normal and Student's T-Error distribution. Entrepreneurship and Sustainability Issues, 7(3), 1580-1596
[25]. Sarfraz, M., Mohsin, M., Naseem, S., & Kumar, A. (2021). Modeling the relationship between carbon emissions and environmental sustainability during COVID-19: a new evidence from asymmetric ARDL cointegration approach. Environment, Development and Sustainability, 1-19. https://doi.org/10.1007/s10668-021-01324-0
[28]. Velavan, T.P., & Meyer, C. G. (2020). The COVID-19 epidemic. Tropical Medicine and International Health, 25(3), 278–280. https://doi.org/10.1111/tmi.13383
[29]. Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., ... & Tan, W. (2020). A novel coronavirus from patients with pneumonia in China, 2019. *New England journal of medicine*.

How to cite this article: Rehman, H. Ur., Shafiq, M., Nabi, F.G., Hassan, S., Baig, S.A. and Manzoor, U. (2021). Analyzing the Epidemiological Outbreak of COVID-19 using Descriptive and Predictive Analytics. *International Journal of Emerging Technologies*, **12**(1): 335–341.