I. J. Education and Management Engineering, **2022, 1, 1-10**
Published Online February 2022 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijeme.2022.01.01

Analyzing the Impact of Vaccination on COVID-19 Confirmed Cases and Deaths in Azerbaijan Using Machine Learning Algorithm

Makrufa Sh. Hajirahimova
Institute of Information Technology of ANAS, City, AZ1141, Azerbaijan
E-mail: hmakrufa@gmail.com

Aybeniz S. Aliyeva
Institute of Information Technology of ANAS, City, AZ1141, Azerbaijan
E-mail: aliyeva.a.s@mail.ru

Received: 14 October 2021; Accepted: 28 November 2021; Published: 08 February 2022

Abstract: For almost two years, the world has been battling a global trouble- the COVID-19 pandemic. The disease, which has spread to about 225 countries around the world, has devastated the healthcare system of even the most developed countries. Governments have found the only way out is to impose a strict quarantine regime and state of emergency. Scientists immediately began testing the vaccine. Vaccination would still be the only savior of the planet's inhabitants.Because many of such pandemic infections have exactly been prevented thanks to vaccines in the past. Although the reduction in the number of infections after strict quarantine measures allowed the restrictions to be eased, the next wave was starting soon. This made it necessary the preparation of the vaccine as soon as possible. At the end of last year, the expected news came. Thus, in December 2020, the vaccination process has been launched in a number of countries. Azerbaijan is also one of the first countries to join the vaccination. The vaccination process, which began on January 18, 2021 continues, provided that 4 types of vaccines are available to the population. As a result of vaccination, the epidemiological situation in Azerbaijan is under control, as in many countries. In this article has been attempted to find a correlation between vaccination and COVID-19-confirmed cases and deaths. For this purpose, the k-means cluster-based machine learning method has been used in the Azerbaijan data collection obtained from the GitHub repository of the Center for Systems Science and Engineering at Johns Hopkins University. This research can benefit governments, stakeholders, and relevant institutions in the health care sector in monitor the vaccination process and more detally assess the epidemiological situation , and make important decisions to control and manage the spread of the disease.

Index Terms: Coronavirus, SARS-CoV-2, COVID-19 pandemic, machine learning, k-means clustering.

1. Introduction

The entire world is currently experiencing the COVID-19 pandemic. COVID-19 is the disease caused by the recently discovered SARS-CoV-2, the coronavirus that first identified in December 2019 in Wuhan, China. The virus quickly began to spread rapidly span around the world, caused millions of people to become infected and die. The COVID-19 became a worldwide public health emergency and a pandemic was declared by the World Health Organization on March 11, 2020 [1, 2].

For confining the dissemination of the virus governments have found the only way out is to impose a strict quarantine regime and state of emergency. Besides, several attempts were proposed to control the pandemic spreading by non-pharmacological interventions like social distancing and the use of face masks. The COVID-19 pandemic has caused large economic losses as a result of the strict measures taken to combat the virus [3, 4]. In addition, the lockdown applied to prevent the spread of the pandemic has affected people's physical health, economic problems, job losses, and etc. also has a negative effect on the relationship with family and spouse. Lockdown has a negative impact on the physical and mental health of the population, causing an increase in negative emotions, stress levels and anxiety of people due to fear of contracting COVID-19 and death due to this [5].

But many people believe in social distancing as an effective measure to break the chain of COVID-19 infection and consider lockdown as the best possible strategy to flatten the curve of infections. Also in some studies [6-9] are noted a large impact of lockdown on the spread of covid-19 virus and its mortality rate.

However, there is a strong consensus globally that a vaccination is one of the most effective and cost effective methods of combating infectious diseases. The COVID-19 vaccine may be the most effective method of controlling infectious diseases today. Because vaccines have helped reduce morbidity and mortality from many diseases over time, they have saved humankind from infectious diseases in the last century. An unprecedented research effort of doctors, scientists, and researchers has resulted in a rapid development of COVİD-19 vaccine candidates and initiation of trials. As known, the process of developing a vaccine typically takes a decade. Also, no effective vaccine has yet been developed almost forty years after the discovery of HIV [10]. As a result, the COVID-19 vaccination comes with new inquietudes related to the relatively short period of time needed for the vaccine development and decreasing the general public's confidence towards vaccination. However, some COVID-19 vaccines have already been approved for use and vaccination programs are being successfully implemented.

Despite the preventive measures taken against this virus in Azerbaijan since the beginning of 2020, thousands of people soon became infected and suffered from this virus. In Azerbaijan, from 3 January 2020 to 27 September 2021, there have been 481401 confirmed cases of COVID-19 with 6476 deaths [11].

Azerbaijan is one of the first in the world that has launched vaccination people. Vaccination against COVID-19 commenced in Azerbaijan on 18th of January 2021, initially using "CoronaVac" vaccine developed by the Chinese Sinovac Biotech. The "Vaxzevria" produced by the British-Swedish AstraZeneca, the Russian "SPUTNIC V" vaccine and Pfizer-BioNTech vaccine was then added to the programme. On May 10, 2021, Azerbaijan has started mass vaccination of all citizens over 18 years old [11, 12]. The main purpose of the vaccination is to prevent the spread of COVID-19, which is currently widespread, by vaccination, thereby reducing morbidity and mortality. As of September 27th, 8.350.575 citizens (81,68% of the entire population) have received the vaccine, of which 4.713.775 (46,39% of the entire population) received the first dose of the vaccine; and 3.636.800 (35,57% of the entire population) received the second dose of the vaccine [13].

In the last hundred years, vaccines have been helped reduce the damage and mortality of many diseases. Our local experts associated the serious reduction in the number of COVID-19 infections and the positive change in the overall epidemiological situation in recent months with carrying out of effective measures, including mass vaccination in the country. The objective of this experiment is analyzing the effect of vaccination on the COVID-19 cases including confirmed cases and deaths, taking into account closures in the country.

2. Literature Review

Below are some studies that analyzed the effect of the vaccination on COVID-19 cases are discussed. Afterwards, a short literature review regarding k-means clustering analysis is conducted in the context of the COVID-19 pandemic.

2.1. Recent studies analyzing the effect of vaccination on COVID-19 cases

Kiyoshi F. Fukutani et al. [14] used the CaVaCo (Cases, Vaccinations and COVID-19) tool to access the COVID-19 database of each countries in real time to assess the impact of COVID-19 vaccination. This tool allows authors to extract the COVID-19 cases, deaths, and vaccination data from all countries, and to compare and relate countries' vaccination coverage with other parameters. The tool is developed in R, and it has the ability to automatically download and standardize data. As a result, it is possible to perform correlation between the number of daily vaccinations, the number of new deaths and the number of tests using the Spearman correlation. Nevertheless, this study has some limitations. This analysis uses quantitative measures and cannot reflect all national behavior or public policy. In addition, the current analysis cannot handle or correct numeric trends or external interference. This study showed that for a more detailed assessment of various epidemiological situations it is necessary to carry out individual analysis for specific countries.

In [15], Mohamed Hussein and colleagues used the Pearson correlation to assess the effects of the previous BCG vaccine on epidemic outcomes, severe critical infections, and mortality cases in more than 180 countries most affected by the pandemic. The results of the study showed that the old BCG vaccine had the ability to prevent severe critical SARS-CoV2 pneumonia and relatively reduce mortality. However, this study has some limitations. First, the rapidly changing nature of the data can influence research results. Secondly, most studies use open data sources of obtaining COVID-19-related information. Another limitation may be due to some problems such as limited testing and reporting for COVID-19 in many countries, which it could also affect the study's results. Another limitation may be the limited scope of COVID-19 testing and reporting in many countries, which may affect the results of the study. Another limitation of the study is the non-stratification of populations by age and co-morbidities, which can be confusing in terms of mortality from COVID-19.

Because both lockdown and vaccination programs affect COVID-19 infections and deaths, it is difficult to assess the impact of each intervention individually. Researchers from British Public Health (PHE) and the University of Warwick [16] used statistical calculations to estimate the impact of the vaccination program on the number of deaths in the UK. Two approaches have been used in the research. In the first approach, the daily effect of the vaccination on mortality was assessed based on the vaccine's efficacy against deaths, taking into account the coverage of the vaccine. In the second approach, vaccinated and non-vaccinated simulations were compared using a dynamics of age - structured model. The results of the study showed that the COVID-19 vaccination program already has a significant impact on COVID-19 disease outside the national quarantine regime.

In [17] has been analyzed the opinions of Twitter users about the vaccination process against the COVID-19 pandemic using BERT (Bidirectional Encoder Representations from Transformers) machine learning technique. The proposed approach has classified tweets into three main classes: in favor, against, and neutral regarding COVID-19 vaccination. During the analysis, it was noticed that the majority of tweets were in a neutral position, and the number of positive tweets exceeded the number of against tweets. This showed that the public is more interested in the vaccination process. The limitations of the study are related to the selected classification algorithms, the chosen dataset, which only includes tweets retrieved over a given period of time, written in English.

Alagoz et al. [18] used an agent-based simulation model to estimate the effect of vaccination on the number of COVID-19 cases in urban communities. The authors found that the vaccination would significantly reduce the number of COVID-19 cases, but the rate of decline in COVID-19 cases varied from region to region, and that nonpharmaceutical interventions (NPIs) had a significant impact on the number of cases. There are a number of limitations to this research. The model does not use the full calibration procedure, which is usually used in simulation, nor does the model take into account age-related differences in the application and effectiveness of the vaccine.

2.2. K-means clustering analysis on COVID-19 data

Zubair et al. [19] proposed an effective clustering method to determine helathcare quality clusters of countries utilizing the COVID-19 datasets. It is known that the k-means clustering algorithm is usually used to create clusters based on similarity. Their proposed method determines the optimal initial centers of the k-means clustering algorithm unlike the previous clustering algorithm. The method also reduces the number of iterations and execution time.

In [20] k-means clustering was used for analysis his results based on the presence of a correlation between the temperature and the three cases of COVID-19 disease like suspected, confirmed and death cases. The datasets for this study were gathered from the WHO for various regions of China. The novelty of this work lies in the fact that authors have included the temperature parameter in the original dataset and further investigate the trends. When authors used k-means clustering, they discovered that the clustering of 30 classes had happened, and the key result was that temperature was not only a factor in the distribution of the COVID-19 pandemic. They proposed that other causes could have played an effective role in the spread of COVID-19 by analyzing the impact of temperatures in the three cases of the disease (confirmed, suspected, and death).

Sheikh Abdullah et al. [21] used k-means clustering and correlation matrix to analyze COVID- 19 risk in pandemic countries. They used k-means clustering to search hidden or unknown clusters in many countries infected with COVID-19, as well as a correlation matrix to identify relationships between the number of attributes.

A study is conducted by Rizvi et al [22] is used a k-means clustering and correlation matrix to analyze the factors influencing the rapid spread of COVID-19 in some countries (in 79 countries). Pearson Product Moment Correlation Analysis is used to identify relationships between of these attributes like health system indicators (PM2.5 exposure, air quality, sanitation score, drinking water score and so on.), disease prevalence (tuberculosis, cardiovascular disease, respiratory infections, asthma so on.) and socio-economic factors (GDP per capita, health expenditure per capita, alcohol consumption, smoking prevalence and so on.) with death cases and confirmed cases to learn about the relationship of these factors with the spread of COVID-9. The k-means clustering algorithm is used to cluster the countries taking into account all the above factors. The results of the experiments showed that the prevalence of the disease is strongly related to COVID-19, while indicators of environmental health are weakly related with COVID-19.

In [23], has been used k-means clustering and RapidMiner tool to analyze situation of COVID-19 cases and deaths in Southeast Asia countries. Southeast Asia countries were grouped of confirmed and death cases of COVID-19 by 3 clusters, that is: high, medium cluster, and low cluster. From the results of clustering, four countries, like Indonesia, Malaysia, Philippines, and Thailand are in the high cluster, where mortality and confirmed cases were highest.

In [24, 25], has been analyzed COVID-19 pandemic risk in provinces of Indonesia based on coronavirus disease data. Provinces in Indonesia were grouped based on the data of confirmed, death, and recovered cases of COVID-19 using k-means clustering method. The clustering results provide the government with information to make policies to against the spread of COVID-19 pandemic.

3. Methodology

This section presents the methodology for analyzing the COVID-19 dataset. The data analysis was conducted in a Python programming environment using the scikit-learn library. The methodology includes several stages. The first step is to collect the COVID-19 vaccination dataset.

3.1. Data Collection

The experiments have been used a dataset, which data were retrieved from the GitHub repository of the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University [26]. We extracted dataset specifying location as Azerbaijan with starting from 1 April 2021 to 27 September 2021. This dataset includes information the number of confirmed cases, death cases, and vaccination and like this all information of coronavirus. We added additional attribute in the data set, type of lockdown given which have been recorded from the Task Force or Operative Headquarter under Prime-Minister website [8]. The reason of appending the new variable because lockdown is considered one of the against factors of spreading the coronavirus. After appending lockdown fields, the results could be more promising for researching the impact of vaccination on COVID-19 cases. The dataset has all the attributes as shown in Table 1. An example record from a dataset is shown in Table 2.

Table 1. Description of the dataset.

№	Variable	Description	Data Type
1.	date	Date of observation	DateTime , Timestamp
2.	new_cases	The number of new COVID-19 confirmed cases per day	Number
3.	new_deaths	The number of new COVID-19 deaths per day	Number
4.	total_vaccinations	Total number of COVID-19 vaccination doses administered	Number
5.	people_vaccinated	The number of people who have received a first dose COVID-19 vaccination	Number
6.	people_fully_vaccinated	The number of people who have received both doses COVID-19 vaccination	Number
7.	people_vaccinated_per_hundred	Total number of people who received at least one vaccine dose per 100 people in the total population	Number
8.	people_fully_vaccinated_ per_hundred	Total number of people who received both doses per 100 people in the total population	Number
9.	type_lockdown	Type of lockdown (Full, Partial or None)	Char

3.2. K-means Clustering Algorithm

Machine learning algorithms are divided into two main learning approaches, like supervised learning and unsupervised learning. Unlike supervised learning, there is no target variable in unsupervised learning. There are only input variables in the dataset that describe the data. K-means is one of the most widely used unsupervised clustering algorithms. K-means clustering is used to find groups of observations (clusters) that share similar characteristics within the unlabeled dataset. Each cluster is represented by the cluster's center, also known as the centroid. In clustering, the task is to split the dataset into multiple groups in such a way that the data points in the same groups are closest to its centroid. So that, each data point is assigned to its nearest centroid, which is based on the squared Euclidean distance. In practice, the distances used in clustering often do not characterize spatial distances. The k-means clustering algorithm inputs are the number of clusters and starting points (the data set). After appointing all data points to the appropriate clusters, the next step is to calculate the new cluster centroid values. This process is iterated until the centroids remain unchanged [27, 28].

Table 2.An example record in the COVID-19 dataset

	date	new_cases	new_deaths	people_vacc	people_fully_vacc	total_vacc_per_hundred	people_vacc_per_hundred	fully_vacc_per_hundred	type_lockdown
0	2021-04-01	2248	26	558485	297878	8.38	5.46	2.91	full
1	2021-04-02	2361	24	569994	319039	8.70	5.58	3.12	full
2	2021-04-03	1852	31	582463	342350	9.05	5.70	3.35	full
3	2021-04-04	2561	29	590914	356831	9.27	5.78	3.49	full
4	2021-04-05	1099	34	591012	356935	9.27	5.78	3.49	full

3.3. Determining the Number of Clusters

The next step is to determine the optimal number of clusters к. There are several methods for estimating the optimal number of clusters k. Here we use the Elbow method for determining the optimal number of clusters k. As shown in Fig. 1, the elbow is at k=3 indicating the optimal value k for this dataset is 3.

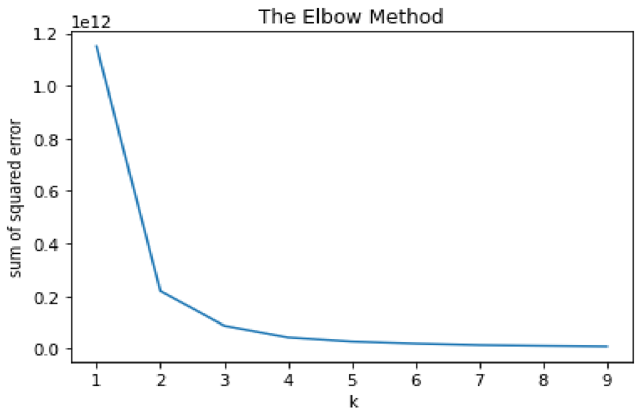


Fig.1. Number of clusters

4. Results and Discussion

In this section, have been presented the results of clustering for COVID-19 data set and analyses each cluster on COVID-19 deaths and confirmed cases. The analysis is made on fully vaccination and death, confirmed cases. These analyses have been used data from two months after COVID-19 vaccination in Azerbaijan. As we have mentioned before, that there have been some closures (lockdowns) during vaccination in the country. Therefore, the effect of vaccination was evaluated taking into account the incidence of infection cases and death during closure (full, partial and none).

4.1. Cluster analysis

Results of k-means clustering are graphically presented in Fig. 2. From the k-means clustering technique 3 clusters (cluster 1 (ful), cluster 2 (partial) and cluster 3 (none)) are found as shown in Fig 2. The elements of cluster 1 are indicated by red points, the elements of cluster 2 are indicated by blue points, and the elements of cluster 3 are indicated by green points and centroids are indicated by red points as shown in fig. 2.

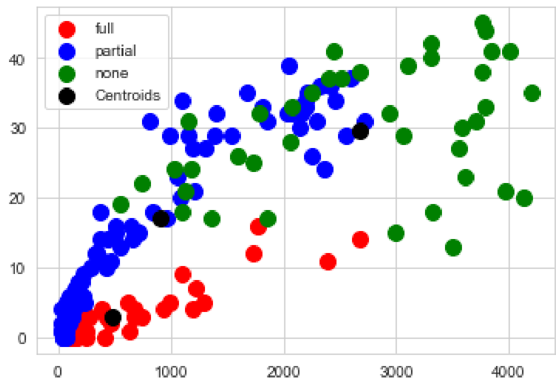


Fig 2. Results of k-means clustering

Cluster 1 (indexed with 0) consist of 89 elements, cluster 2 (indexed with 2) comprises of 54 elements and cluster 3 (indexed with 1) comprises of 27 elements as shown in Table 2.

Table 2. Cluster Results on COVID-19 Dataset

	date	new_cases	new_deaths	people_vacc	people_fully_vacc	people_vacc_ per_hundred	fully_vacc_ per_hundred	cluster
0	2021-04-01	2248	26	558485	297878	5.46	2.91	0
1	2021-04-02	2361	24	569994	319039	5.58	3.12	0
2	2021-04-03	1852	31	582463	342350	5.70	3.35	0
3	2021-04-04	2561	29	590914	356831	5.78	3.49	0
4	2021-04-05	1099	34	591012	356935	5.78	3.49	0
...	...	...	...	...	...	...	...	...
84	2021-06-27	40	1	2138684	1209743	20.92	11.83	0
85	2021-06-28	38	2	2139001	1210034	20.92	11.84	0
86	2021-06-29	87	3	2174186	1248055	21.27	12.21	0

87	2021-06-30	86	1	2214990	1296229	21.67	12.68	0
88	2021-07-01	75	1	2256816	1340292	22.08	13.11	0
89	2021-07-02	101	2	2290808	1379921	22.41	13.50	2
90	2021-07-03	95	0	2321863	1418584	22.71	13.88	2
91	2021-07-04	99	1	2345077	1447687	22.94	14.16	2
92	2021-07-05	62	0	2345371	1448162	22.94	14.17	2
93	2021-07-06	78	0	2371796	1483659	23.20	14.51	2
…	…	…	..	…	…	…	…	…
138	2021-08-22	3069	29	3713143	2457233	36.32	24.04	2
139	2021-08-23	2945	32	3713422	2457302	36.32	24.04	2
140	2021-08-24	3583	30	3772533	2477140	36.90	24.23	2
141	2021-08-25	3794	33	3830650	2504655	37.47	24.50	2
142	2021-08-26	3714	31	3884713	2534824	38.00	24.79	2
143	2021-08-27	4203	35	3929404	2564353	38.44	25.08	1
144	2021-08-29	4010	41	4042976	2644830	39.55	25.87	1
145	2021-08-30	3107	39	4043744	2645287	39.55	25.87	1
146	2021-08-31	3788	44	4113716	2692024	40.24	26.33	1
147	2021-09-01	3845	41	4179684	2731755	40.88	26.72	1
…	...	…	…	…	…	…	…	…
165	2021-09-23	1178	24	4668637	3500836	45.67	34.24	1
166	2021-09-24	1128	21	4683453	3544276	45.81	34.67	1
167	2021-09-25	1099	18	4700096	3593494	45.97	35.15	1
168	2021-09-26	1029	24	4713750	3636195	46.11	35.57	1
169	2021-09-27	558	19	4713775	3636800	46.11	35.57	1

4.1.1. Clustering Results for COVID-19 Confirmed Cases

The cluster results of data points are graphically represented in Fig. 3, where X-axis represents fully vaccination and Y-axis represents the confirmed cases. As shown in Fig. 3, the number of clusters for confirmed cases is 3.

Cluster 1(full lockdown - blue points) consist of 89 elements as shown in Table 2. Cluster 1 (full lockdown) results are graphically represented (indicated with 0) in Table 2. From the analysis of clustering results, shown in Fig.3, we observe a decrease in the number of confirmed cases with an increase in the number of fully vaccinated. In Fig.4 (a) it can also be seen decreasing of confirmed cases with increasing of fully vaccination. As we have seen in Table 2. , fully vaccination per hundred hasn't the high value i.e., around 13.11%. If we relate the cluster results with fully vaccination percentage value as shown in Fig. 3, the trend seems there is no as such direct relationship between the at such low rates of vaccination percentage value and confirmed COVID-19 cases. Some exception cases such as when the cluster value for fully vaccination per hundred reaches to ≥ 35%° then the number confirmed cases decrease 4203 to 558 as shown in Table 2. In this case, the decrease in the incidence of COVID-19 infection may be due to closures during vaccination. As mentioned above, there were certain lockdowns in the country from March 17, 2021 to June 1, 2021.

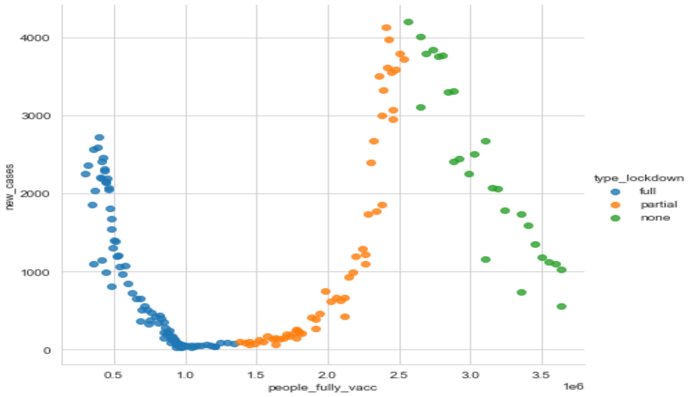
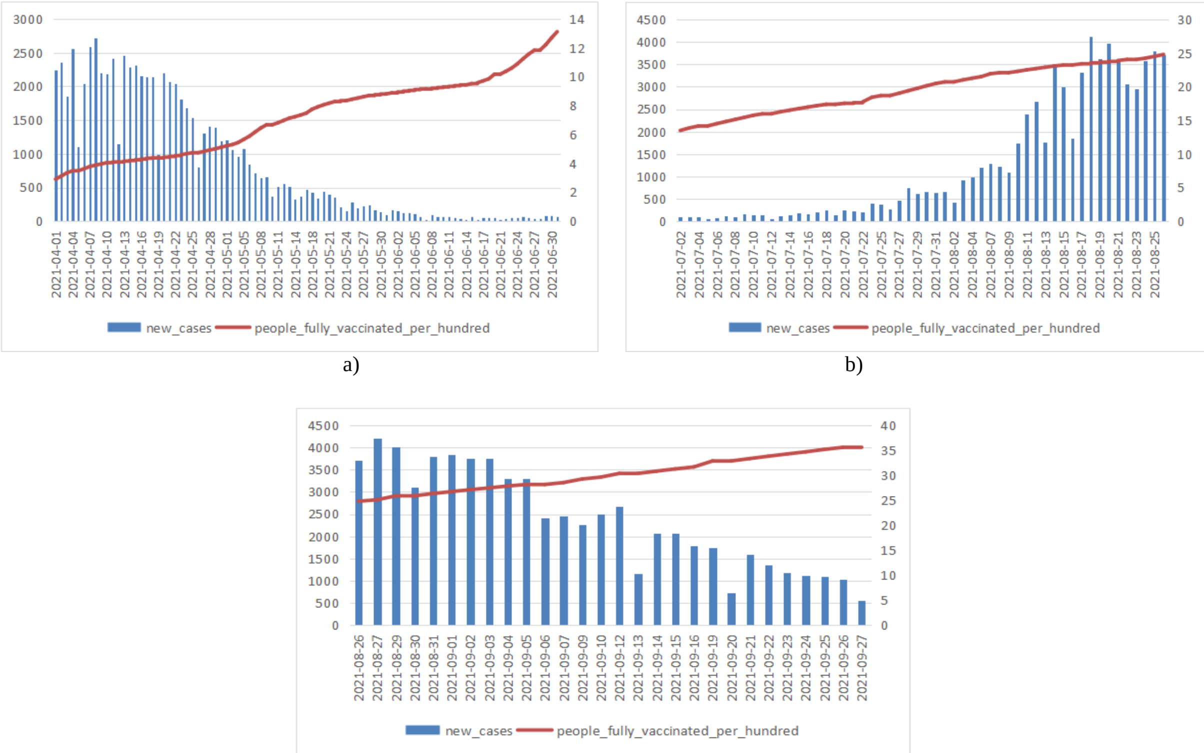


Fig. 3. Cluster results on the effect of fully vaccination on confirmed cases by COVID-19

Cluster 2 (partial lockdown) consist of 54 elements (indicated with 2) as shown in Table 2. Cluster 2 (partial lockdown) results are graphically represented (orange points) in Fig. 3. From the analysis of clustering results, shown in Fig.3, we observe a increase in the number of confirmed cases with an increase in the number of fully vaccinated. In Fig.4 (b) it can also be seen increasing of confirmed cases with increasing of fully vaccination percentage. As we have seen in Table 2, the number confirmed cases increase 78 to 3794. The lockdowns have been removed in the country; only slight restrictions remained from June 1, 2021. And the cluster value for fully vaccination percentage is 24.79% (number of fully vaccinations - 2.534.824). The trend seems, there is no high correlation between a small percentage of fully vaccination percentage and confirmed COVID-19 cases.



c)

Fig. 4. The effect of fully vaccination on confirmed cases.

Cluster 3 (none lockdown) include 27 elements (indicated with 1) as shown in Table 2. Cluster 3 (green points) results are graphically represented in Fig. 3. From the analysis of clustering results, shown in Fig.3, we observe a decrease in the number of confirmed cases with an increase in the number of fully vaccinated. In Fig.4 (c) it can also be seen decreasing of confirmed cases with increasing of fully vaccination percentage. As we have seen in Table 2, when the cluster value for fully vaccination percentage reaches to ≥ 35% (number of vaccination- 3.636.800), then the number confirmed cases decrease from 4203 to 558. As mentioned above, the lockdowns have been removed in the country from July 27, 2021. Despite the removed of closures the COVID-19 vaccination programme having a significant impact on COVID-19 disease. As can be seen, there is direct relationship between the high percentage of fully vaccination (≥ 35%) and confirmed COVID-19 cases.

4.1.2. Clustering Results for COVID-19 Deaths Cases

The clustered death cases are graphically represented in Fig. 5, where X-axis represents death cases and Y-axis represents the fully vaccinations. As shown in Fig. 5 there are 3 clusters for death cases: the elements of cluster 1 (full) are indicated by blue points, the elements of cluster 2 (partial) are indicated by orange points, and the elements of cluster 3 (none) are indicated by green points .

Cluster results manipulated in Fig. 5, same trend has been detected as we have seen in Fig. 3. Here, too, we observe a decrease in the number of death cases with an increase in the number of fully vaccinated. Note that, depending on the lockdowns and fully vaccination percentage can be seen that death cases may increase or decrease in some cases as shown in Fig. 5.

As we have seen in Fig.6 (a) that clusters 1, the decreasing of death cases with increasing of fully vaccination percentage. For comparison, it can be noted that if in 2 months after the start of the vaccination (second dose), i.e. in April, the number of daily mortality case ranged from 25 to 40, in June this figure was 0-8. In this case, the decreasing of COVID-19 deaths may be due to closure during vaccination, which it was carried out in the country from March to June 1, 2021. Because, as can be seen in Fig. 8 (a) and in Table 2, the vaccination rate in March-June is not high enough.

Cluster 2 (orange points) results are graphically represented in Fig. 5. From the analysis of clustering results, shown in Fig. 5, we observe an increase in the number of deaths with an increase in the number of fully vaccinated. In Fig.6 (b) it can also be seen increasing of death cases with increasing of fully vaccination percentage. Despite the cluster value for fully vaccination percentage is 24.79% (number of fully vaccinations – 2.534.824), increasing the number of deaths (from 1 (or 0) to 33) shown in Table 2. Note that, the lockdowns have been removed in the country, only slight restrictions remained from June 1, 2021. As can be seen, there is no strong association between the as such low rates of vaccination percentage value and COVID-19 death cases.

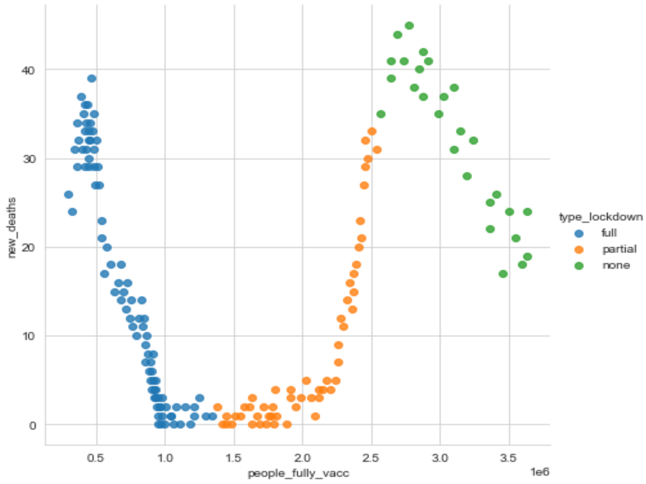


Fig. 5. Cluster results on the effect of fully vaccination on death cases by COVID-19

Cluster 3 (blue points) results are graphically represented in Fig. 5. From the analysis of clustering results, shown in Fig.5, we observe a decrease in the number of death cases with an increase in the number of fully vaccinated. In Fig.6 (c) it can also be seen decreasing of deaths with increasing of fully vaccination percentage. As we have seen in Table 2, when the cluster value for fully vaccination percentage reaches to ≥ 35%, then the number deaths decrease from 44 to 18 (in case of removing restrictions).

Fig. 6. The effect of fully vaccination on reducing of death cases by COVID-19.

The experimental results showed that the COVID-19 vaccination programme is already having a significant impact on severe COVID-19 disease beyond the effect of the national lockdown. If we interrelate the cluster results with fully vaccination percentage value as shown in Fig. 4 and Fig.6, we can see there is have as such direct relationship between the at such highest rates of vaccination percentage value and COVID-19 confirmed cases and death cases. It should be noted, that the potential precautions can slow the COVID-19 infection spread also, when COVID-19 vaccination rate is low.

5. Conclusion

In this study, has been attempted to estimate the association between vaccination and COVID-19 confirmed cases and death cases. For this purpose, the k-means cluster-based machine learning method has been employed on the data set of Azerbaijan, the analysis is made on fully vaccination and death, confirmed cases.

These analyses have been used data from two months after COVID-19 vaccination in Azerbaijan. As we have mentioned before, that there have been some closures (lockdowns) during vaccination in the country. Therefore, the effect of vaccination was evaluated taking into account the incidence of infection and death during closure (full, partial and none). We have implemented the unsupervised k-means clustering and we find that k=3 (applying the elbow method) is the optimal number of cluster for our dataset. We have find that the model has low inertia of 52.096. We have achieved a classification accuracy of 68% with k=3 by our model and 116 out of 170 samples were correctly labeled. We think so this is a good model fit to the dataset.

In summary, while investigating the impact of vaccination in confirmed and death cases, we found that fully vaccination is the considerable and important factor for the effect on COVID-19 cases. We can say that other attributes like lockdowns are also playing the role in the COVID-19 pandemic which slowed the spreading infection. Before getting vaccinated, no potential precaution has been seen except the lock-down, which has been seen in Azerbaijan against the COVID-19 pandemic. However, shortly after the lockdowns were removed, we saw a rapid increase in the number of confirmed cases. After increasing number of vaccination (or increasing of fully vaccination percentage), we have been witnessing a different view. The experimental results provide further evidence that the COVID-19 vaccination programme is already having a significant impact on severe COVID-19 disease beyond the effect of the national lockdown. There has been no significant increase in the number of people infected with COVID-19, despite the restoration of action of many workplaces, large shopping centers, subways and restaurants in the country. Vaccination is a powerful armor against the COVID-19 pandemic. Furthermore, the potential precautions can slow the COVID-19 infection spread, when COVID-19 vaccination rate is low. The results of research can be utilized by policy makers and relevant institutions in the health care sector to make better decisions to control the pandemic, in monitor the vaccination process and more detally assess the epidemiological situation. Also, these results offer a strong foundation for conducting research on reliable prognostic estimations and analysis of vaccines effectiveness to prevent infection.

References

[1] E. Mathieu, H. Ritchie, E. Ortiz-Ospina , M. Roser and et al,," A global database of COVID-19 vaccinations ", Nature Human Behaviour, 2021, www.nathumbehav.com

[2] The World Health Organization, https://covid19.who.int/info

[3] Q. Li, X. Guan, P.Wu, X. Wan, "Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia," New England Journal of Medicine, vol. 382, no. 13, pp. 1-9, 2020, DOI:10.1056/NEJMoa2001316

[4] A. H. Mohammed Hassan, A. A. Mohammed Qasem, W. F. Mohammed Abdalla, O. H. Elhassan, "Visualization & Prediction of COVID-19 Future Outbreak by Using Machine Learning", International Journal of Information Technology and Computer Science(IJITCS), vol.13, no.3, pp.16-32, 2021. DOI: 10.5815/ijitcs.2021.03.02

[5] M.W. Hussain, T. Mirza, M. M. Hassan. " Impact of COVID-19 Pandemic on the Human Behavior ", International Journal of Education and Management Engineering (IJEME), vol.10, no.5, pp.35-61, 2020. DOI: 10.5815/ijeme.2020.05.05

[6] A.Tobías, "Evaluation of the lockdowns for the SARS-CoV-2 epidemic in Italy and Spain after one month follow up", Science of the Total Environment. vol. 725, 10 July 2020, https://doi.org/10.1016/j.scitotenv.2020.138539

[7] D.Ibarra-Vega, " Lockdown, one, two, none, or smart. Modeling containing covid-19 infection. A conceptual model", Science of the Total Environment. vol. 730, 15 August 2020, https://doi.org/10.1016/j.scitotenv.2020.138917

[8] B.Ambikapathy, K. Krishnamurthy, "Mathematical modeling to assess the impact of lockdown on COVID-19 transmission in India: Model development and validation", JMIR Public Health Surveill., vol. 6, no.2., 2020.

[9] C.Violato, E.M,Violato, E.M.Violato, "Impact of the stringency of lockdown measures on covid-19: A theoretical model of a pandemic", PLoS ONE , vol.16, no. 10, 2021, https://doi.org/10.1371/journal.pone.0258205

[10] T. T. Le, Z. Andreadakis, A. Kumar, R. G. Román, S. Tollefsen, M. Saville, and S. Mayhew, ''The COVID-19 vaccine development landscape,'' Nature Reviews Drug Discovery, vol. 19, no. 5, pp. 305–306, 2020, doi: 10.1038/d41573-020-00073-5.

[11] Task Force or Operative Headquarter under Prime-Minister of the Republic of Azerbaijan, https://nk.gov.az/

[12] AZERTAC - Azerbaijan State News Agency, azertag.az

[13] https://koronavirusinfo.az/az

[14] F. Fukutani, M.L. Barreto, B.B. Andrade and A.. L. Queiroz, "Correlation Between SARS-Cov-2 Vaccination, COVID-19 Incidence and Mortality: Tracking the Effect of Vaccination on Population Protection in Real Time", Frontiers in Genetics, vol 12, pp.1-4, 2021.

[15] A. R. M. Hussein, M. R. Salem et al.,"Correlation between COVID-19 case fatality rate and percentage of BCG vaccination: is it true the vaccine is protective?", The Egyptian Journal of Bronchology, vol. 14, no. 25, 2020.

[16] N. Andrews, M. J. Keeling, J. Stowe, Sh. Ismael, L. Coughlan, H. Allen, M. Ramsay, J. L. Bernal,"Impact of COVID-19 vaccines on mortality in England: December 2020 to February 2021", Public Health England, March 2021, www.gov.uk/phe

[17] L.-Adrian Cotfas, C. Delcea, I. Roxin, C.Ioanăş, D.S. Gherai and F. Tajario, “The longest month: analyzing covid-19 vaccination opinions dynamics from tweets in the month following the first vaccine announcement”, IEEE Access, vol. 9, pp. 33203-33223, 2021, doi:10.1109/ACCESS.2021.3059821
[18] Alagoz O, Sethi AK, Patterson BW, Churpek M, Alhanaee G, Scaria E, et al., “The impact of vaccination to control COVID-19 burden in the United States: A simulation modeling approach”, PLoS ONE, vol. 16, no. 7, pp. 1-12, 2021, https://doi.org/10.1371/journal.pone.0254456
[19] Md. Zubair, A. Iqbal, A. Shil, E. Haque, M. M. Hoque, I.l H. Sarker, “An Efficient K-means Clustering Algorithm for Analysing COVID-19”, arXiv:2101.03140v1 [cs.LG] 21 Dec 2020.
[20] K. Siddiqui, R.Morales-Menendez, P.K.Gupta and et al., “Correlation Between Temperature and COVID-19 (Suspected, Confirmed and Death)Cases based on Machine Learning Analysis,” J Pure Appl Microbiol, vol. 14(suppl 1), pp. 1017-1024, May 2020.
[21] S. H. Sheikh Abdullah et al., “Clustering and Correlation Methods for Predicting Coronavirus COVID-19 Risk Analysis in Pandemic Countries”, 8th International Conference on Cyber and IT Service Management (CITSM), 23-24 Oct. 2020, Pangkal, Indonesia.K.
[22] S.A. Rizvi, M.Umair, M. A. Cheema, “Clustering of Countries for COVID-19 Cases based on Disease Prevalence, Health Systems and Environmental Indicators”, Chaos, Solitons and Fractals 151 pp. 1-10, 2021.
[23] J. Hutagalung, N. L. W. Sri Rahayu Ginantra, G.W. Bhawika, W. G. Suka Parwita , A.Wanto, P. D. Panjaitan, “COVID-19 Cases and Deaths in Southeast Asia Clustering using K-Means Algorithm “, Journal of Physics: Conference Series 1783 (2021) 012027, IOP Publishing, doi:10.1088/1742-6596/1783/1/012027
[24] D.Abdullah, S.Susilo, A.S.Ahmar,R.Rusli and R.Hidayat, “The application of K-Means clustering for province clustering in Indonesia of the risk of the COVID-19 pandemic based on COVID-19 data”, Quality & Quantity, 2021, https://doi.org/10.1007/s11135-021-01176-w
[25] F. Virgantari and Y. E. Faridhan, “K-Means Clustering of COVID-19 Cases in Indonesia’s Provinces”, Iinternational journal of Eengineering and Natural science, vol. 5, no. 2, November 2020.
[26] Data on COVID-19 (coronavirus) by Our World in Data, https://github.com/owid/covid-19-data/tree/master/public/data
[27] Understanding K-means Clustering in Machine Learning (by Dr. Michael J. Garbade), Towards Data Science , Sep 13, 2018, https://towardsdatascience.com/k-means-clustering-in-machine-learning
[28] B. Rawat, S. K. Dwivedi, "Analyzing the Performance of Various Clustering Algorithms", International Journal of Modern Education and Computer Science(IJMECS), Vol.11, No.1, pp. 45-53, 2019.DOI: 10.5815/ijmecs.2019.01.06

Authors’ Profiles

Makrufa Sh. Hajirahimova has been working as a department head at the Institute of Information Technology of Azerbaijan National Academy of Sciences. She teaches at the Training Innovation Center of the institute. She defended the thesis on the “Development of models and algorithms for intelligent management of text document in e-government” at the Institute of Information Technology of Azerbaijan National Academy of Sciences and received PhD in Technical sciences in 2013. In 2016, she was awarded the title of Associate Professor by the Higher Attestation Commission under the President of the Republic of Azerbaijan. She actively involved in the development of the "e-Science" software project under the "E-Azerbaijan" program. Her current research interests include Electron demography, social network security, anomaly detection, Big Data Analytics and machine learning. She is the author of more than 120 papers. More than 35 of her works were published abroad, and several were published in international journals with high impact factor.

Aybeniz S. Aliyeva graduated from Applied Mathematics faculty of Baku State University. After graduating from university she began working Institute of Information Technology of ANAS. At the moment she is a senior researcher at the Institute of Information Technology. Her research interests include Electron demography, Big Data, Big Data Analytics and machine learning. She is the author of more than 60 papers. More than 20 of her works were published in international journals and conferences.

How to cite this paper: Makrufa Sh. Hajirahimova, Aybeniz S. Aliyeva, "Analyzing the Impact of Vaccination on COVID-19 Confirmed Cases and Deaths in Azerbaijan Using Machine Learning Algorithm", International Journal of Education and Management Engineering (IJEME), Vol.12, No.1, pp.1-10, 2022. DOI: 10.5815/ijeme.2022.01.01