

Medicine Allotment For COVID-19 Patients Using Statistical Data Analysis

LAKSHMI SUMANTH SAKALA¹, YESHWANTH ZAGABATHUNI²

^{1, 2} Student, Vignan's Foundation for Science, Technology and Research

Abstract—Computational intelligence deals with development and application of computational models and simulations and is often coupled with high performance computing to solve complex problems arising in engineering analysis and design as well as natural phenomena. COVID-19 is a coronavirus-induced infectious disease. Most people worldwide got infected with this virus and became mild to moderately ill with respiratory related problems. Most infected individuals who experienced mild to moderate illness/ disease and without hospitalization recovered. Yet older people with underlying medical conditions are more likely to experience severe diseases, such as cardiovascular disease, diabetes, chronic respiratory disease, and cancer. As specific vaccine or treatment for COVID-19 is not yet prescribed, it is a tough task to prescribe a common medicinal procedure. There are many ongoing clinical trials evaluating potential treatments. This work presents an application that allots medicines to the one who tested positive. This proceeds after checking patients' medical data which include BP, diabetes, cancer, alcoholic habits etc. Variations in the patient data originated from various sources with several medical concerns with different specifications is useful in evaluating and allotting proper medical course for COVID-19 patient treatment. Numerous attributes are used in creating the database. Different ages are categorized and the corresponding treatment will be prescribed based on the age category and the medical history of the patient. Missing data can affect the performance of data mining system. This work presents clustering which is a method of unsupervised learning and common technique for statistical data analysis. Various clustering algorithms with test samples are carried out for medicine allotment based on age category, symptoms and medical history to evaluate the respective accuracy score.

Keywords—COVID-19, clustering, Patient, Algorithms, Accuracy, Medicine Allotment, Data Analysis, Statistics

I. INTRODUCTION

COVID-19, also named as 2019-Novel coronavirus (2019-nCoV), is severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). This virus emerged in Wuhan City of China. It was first identified by Dr. Li Wenliang, who was called to be a hero for raising the alarm of the COVID-19 in early days of the pandemic outbreak. On 11th March of 2020, WHO-World Health Organization declared this disease as a pandemic. A global coordinated effort is needed to stop the further spread of the corona virus. The first confirmed case of COVID-19 in china was traced on November 17, 2019. Coronavirus disease is an acute infectious respiratory disease.

Coronavirus that became a pandemic first affected a person on 17th November 2019. It emerged as the first example of "Disease X", a hypothetical disease of humans caused by an unknown infectious agent, which was named to be novel corona virus and designed as (SARS-CoV-2) severe acute respiratory syndrome coronavirus2. Though some details are unknown, the first origin outbreak took place at the animal market in Wuhan, China and China mentioned it as a case of zoonotic spill-over. SARS-CoV-2 appears to have evolved from Bat-CoV isolate RaTG13, which diversified from a common ancestor from which Pangolin-CoVs have also evolved. The admixed Bat-CoV sub-populations indicate that bats serve as reservoirs harboring virus that are responsible for zoonotic spill-overs (A disease which can be transmitted to humans from animals).

II. HUMAN CORONA VIRUS SPREAD AND SYMPTOMS

Human Corona virus types: Scientists have divided corona viruses into four sub – groupings named as alpha , beta , gamma, and delta. Seven viruses that can infect people/humans are 229E (alpha), NL63 (alpha), OC43 (beta), HKU1 (beta), MERS-CoV : a beta virus that causes Middle East Respiratory Syndrome(MERS), SARS-CoV : a beta virus that causes Severe Acute Respiratory Syndrome(SARS), SARS-CoV-2 : which causes COVID-19.

As humans we leave droplets while we sneeze, cough and mitigate when various surfaces are touched with the same hands. Other living beings get infected by touching the same objects or surfaces with their hands and may touch their eyes, nose and mouth before cleaning their hands. The virus enters our body through droplets which are at a size of about 5µm .COVID-19 can also enter into the body through various ways. As hands touch too many surfaces and can quickly pickup viruses. By this the virus is transferred to your face, mainly where the virus can move into your body, which makes you feel unwell. The life spans of corona on different types of materials are:

Material	Life-Span
Air	3 hours
Copper	4 hours
Cardboard	24 hours
Stainless Steel	2-3 days
Polypropylene Plastic	3 days

The researchers compared the lifespans of COVID-19 on different surfaces in a 70 – degree Fahrenheit room at 40% relative humidity. They found that coronavirus lived the longest on stainless steel and polypropylene, a type of plastic used in everything from food – storage, containers, toys, etc. According to group of experts, high temperatures and muggy weather might make the virus less contagious. The exact lifespan of the corona virus on a surface cannot be determined exactly, subway pole, stairwell banister or money. It depends on many factors, including the surrounding temperature, humidity, and type of surface. On an average a corona virus affected patient infects at least

2 persons, which is more dangerous than flu. There is more than just one strain of new corona virus. Researchers have studied changes in corona virus RNA to figure out that how different types of corona viruses are related to each other. Researchers have taken 103 samples of people who were affected by the new corona virus and they also looked at corona viruses from animals. It was clearly found that corona viruses found in humans and animals were different. According to research there are two types of corona viruses, which the researchers called “L” and “S”. Both the types are very similar with a very slight difference. As corona virus keeps on spreading over the world, the virus continuously keeps changing. It is to the experts that they might find new strains of the virus. Change is what viruses do, and it’s impossible to predict how those viruses change at any moment and what happens is a question mark. N95 masks are designed to remove 95% of all the particles that are at least 0.3µm in diameter.

The virus can cause a range of symptoms from mild illness to pneumonia. There is no data that it does not affect people of a certain age group. The latest evidence shows that there are two groups of people at higher risk of suffering from this disease: The people above 60 years and people with medical conditions such as diabetes, chronic respiratory diseases, cardiovascular disease and cancer. It gradually increases the risk of the people aging above 40 years. It’s important for the people from these age ranges to protect themselves from the disease. The people who consume alcohol might be associated with a range of communicable and non-communicable diseases and mental health disorders, which make the person vulnerable to the virus. Consumption of alcohol weakens the immune system of one’s body and it leads to increase health risks. Hence, the people who consume alcohol should minimize their consumption at any time and its necessary to minimize during the corona pandemic. And COVID-19 causes more severe illness than seasonal influenza. Globally many people have built immunity towards different seasonal flu strains. Corona is a new virus to which no one has the answer. That means more number of people are susceptible to infection, and some of them will suffer severe disease and hence everyone need to enhance their immunity.

A. Symptoms

SARS-CoV-2(COVID-19) affects people in many ways. There are many symptoms that affect the people who are affected by COVID-19. Below are a few symptoms listed by COVID-19 patients:

- 1) Most common symptoms: Fever, Dry Cough and Tiredness.
- 2) Less common symptoms: Aches and pains, Sore throat, Diarrhea, Conjunctivitis, Headache, Loss of taste and smell, rash on skin, discoloration of fingers and toes.
- 3) Serious symptoms: Difficulty in breathing or shortness of breath, Chest pain or pressure, Loss of speech or movement.

If the above listed symptoms are identified, immediate medical attention should be requested. If a person is with mild symptoms, it is suggestible to manage their care at home by taking several precautionary and preventive methods. On an average it takes about 5-6 days for someone to find whether he /she is infected or not. Sometimes it may even take 14 days to show the symptoms of COVID-19.

III. COVID-19 MEDICATIONS

The increase in the spread started due to neglecting the virus as everyone were very confident thinking of it as normal cold . Many people migrated to different countries and this was the reason for spread. As people were staying in groups and moving out of home without proper safety, many people were infected. And the virus also spreads due to droplets, so when specific safety is not taken such as using masks, contact with surfaces and placing hands on your face leads to spread of virus into body easily. The spread can be reduced when every precaution is taken when going out and following the lockdown rules. Negligence of the rules leads to effect of COVID-19.

A. Precautionary Measures

Safety Precautions to be taken to not be affected by COVID-19:

- 1) Make sure to stay clean.
- 2) Maintain Social Distancing,

- 3) Masks are necessary to avoid COVID-19 (N95, KN95 etc).
- 4) Make sure to sanitize your hands when made a contact.
- 5) Do not contact with the surroundings.
- 6) You must not place your hands on your face.
- 7) Do not move on different places as it also passes through droplets.
- 8) When returned from different places make sure that, you wash your mask every single time.
- 9) Do not migrate to different places.

IV. APPLICATION DEVELOPMENT FOR ALLOTMENT FOR MEDICINE

A. Methodology

This work presents an application that allots medicines to the one who tested positive. This is made after checking blood pressure, diabetes, alcoholic, any cancer details etc. The crisis of COVID-19 alerted researchers to face the pandemic and also to come up with solutions for present and for the predicted future. Variations in the symptoms of patients show the requirement of different medicine allotment. The issue of world crisis of COVID-19 is addressed and medicine allotment is done to COVID-19 patients with the collected data. Figure 1 represents the process and the cluster evaluation.

B. Database Creation

Data can originate from different sources and needs to be checked before it can be put to use. This can be done by directly importing files that may already be available in .csv or .xls formats. The attributes we used in the dataset are Age, Diabetes i.e., whether it is positive or negative, Diabetes Level (High/ Normal/ Low), BP Level (High/ Normal/ Low), Cancer, (positive/ negative), alcoholic (positive/ negative), medicine and here medicines include many types based on the above attribute data. The following are the number of attributes used in creation of a database.

The data as in Table I is a sample of the data set that had been taken to give a brief idea of the ages and different categories that are taken to give the medicine for the patients tested positive for COVID-19. Mainly there are 5 different types of categories in ages like 0-15, 16-40,41-60,61-80,80+, everyone is given

medicine according to their ages and the extra categories listed above.

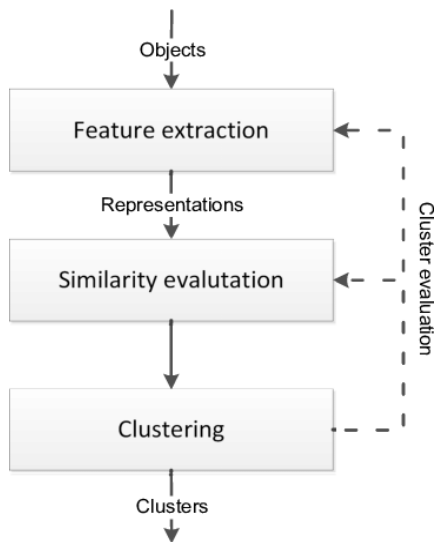


Fig. 1. Flowchart

TABLE I. NUMBER OF ATTRIBUTES USED IN THE CREATION OF A DATASET

C. Data Pre-processing

In any data mining applications datasets may have missing data values. These missing values can get propagated due to lack of communication among the parameters in a collection system. The missing values may affect the performance of a data mining algorithm, and must be handled.

Clustering seems to be the most appropriate type of algorithm to implement proposed solution. The steps to be followed to allot medicines are:

1) You need to take the details of the patient.

2) The details to be considered are:

- Age of the patient
- Detail whether the patient is diabetic or not
- The diabetic level of the patient
- Checking the BP level of the patient
- To know whether the patient is cancer affected or not
- Data whether the patient is alcoholic or not

Based on the considerations of above data, medicine is allotted to specific patient using various algorithms as follows:

3) Canopy

Canopy is a method that is carried out before applying K-Means algorithm.

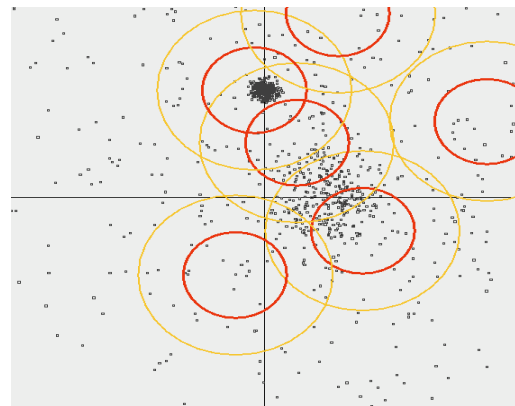


Fig. 2. Canopy

As shown if figure 2, grouping operations on massive data are designed to increase effective complete application of the system given the magnitude of the set of data. The algorithm proceeds using two thresholds T1 i.e the loose distance and T2 i.e, Tight distance. When data is considered, it doesn't give the accurate results.

4) Cobweb

Cobweb is an incremental system for hierarchal conceptual clustering. COBWEB incrementally organizes observations into a trees.

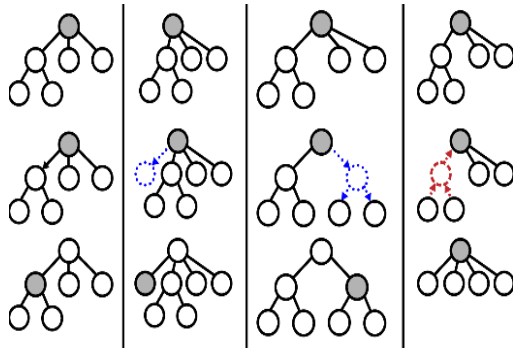


Fig. 3. COBWEB

In such a tree structure each node in the graph a class (definition) and a deterministic term sums up the price ranges of the attributes of items labeled as nodes as depicted in figure 3. When compared to dataset it doesn't give an accurate result so we go to the next clustering algorithm.

5) EM

An alternative to the linear regression assessment in the context of the current parameters is the prediction method. First, the latent variables are calculated, therefore the template is optimised, then those two phases have been repeated before consistency. It is an efficient and general method and most widely used for estimating volume with original value, for example with cluster analysis, as shown in Fig.4.

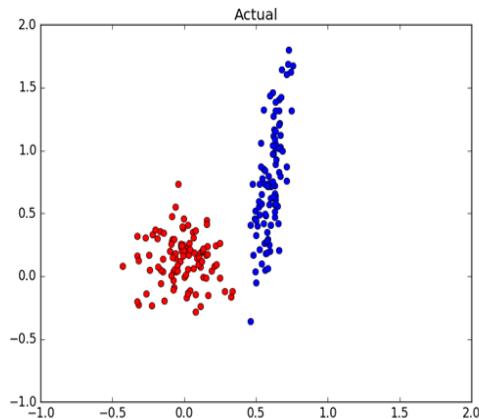


Fig. 4. EM – Expectation – Maximation

6) Farthest First

The furthest first proposed method is a revised k-mean which, as shown in Figure 5, positions each cluster center more closely than the actual cluster center. This must be within the field of data. In certain cases, this massively improves group speed, as fewer changes and reassignments are required. We get a precision of 55

percent whenever this method is based in the current data set.

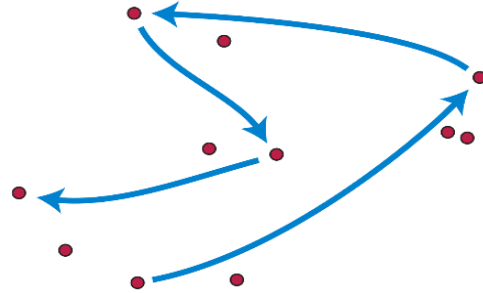


Fig. 5. Farthest First Clustering

*The above image has been taken from wikipedia.

7) Hierarchical Clustering

Hierarchical clustering requires the establishment of clusters with a default structure through start to finish. For instance, all hard disc files and directories are organised as in Figure 6 in a category.

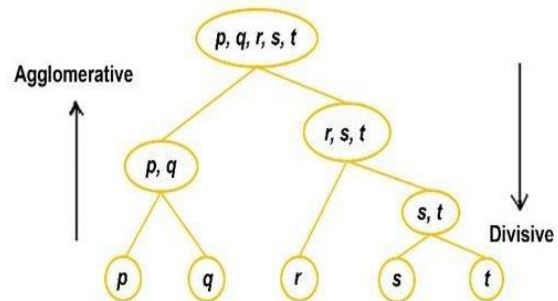


Fig. 6. Hierarchical Clustering

*The above image has been taken from a blog in the kdnuggets website.

There are two types of Hierarchical clustering i.e., Divisive and Agglomerative. When it comes to the execution of the dataset this clustering method gives an accuracy of 42% which is lower than the Farthest First algorithm.

8) Density based Clustering

Quality of the Clustering Algorithm refers to unmonitored optimization techniques which classify different activists in the data on the basis that a cluster within a data space is an adjoining high density area separated by adjacent low-point density regions by adjacent clusters as shown in Figure 7. The data points are generally called noise/outliers in the separating low-point areas. This results in a precision of 58 percent.

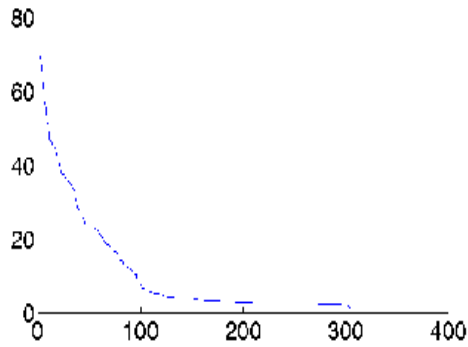


Fig. 7. Density based Clustering

9) Simple K-Means

Simple K-means clustering algorithm attempts to cluster the sample into predetermined non-overlapping subgroups (Clusters). It aims to make the intra-cluster data sets as reasonably practicable while at the same time preserving the clusters as diverse (inter-class). It adds data points to a cluster such that the distance between a data point and a cluster center is smallest. As the variance between clusters is less the (similar) data sets are diverse in the same cluster as in figure 8. This algorithm gives an accurate result when compared to all the clusters with an accuracy of 65 percent.

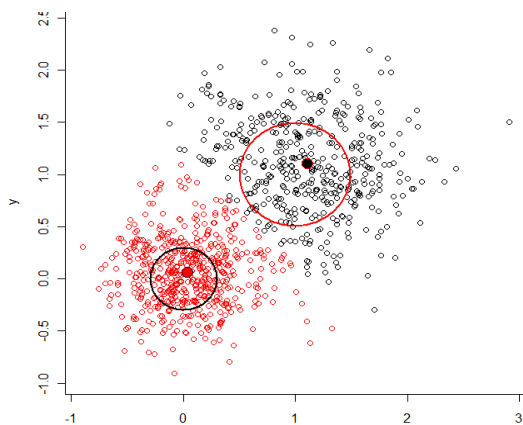


Fig. 8. Simple K-Means Clustering

TABLE II. NUMBER OF ATTRIBUTES USED IN THE CREATION OF A DATABASE

Displaying various models and respective accuracy score

Clusters	Number of Testing Samples	Accuracy
----------	---------------------------	----------

Canopy	216	25 %
Cobweb	216	5 %
EM	216	38 %
Farthest First	216	55 %
Hierarchical	216	42 %
Density based	216	58 %
Simple K-means	216	65 %

The above table II illustrates the clusters used and the number of test samples with the accuracies of the clusters.

10) Pie Chart

The pie chart below shows the accuracy of all the clusters from higher to lower accuracy in which, Simple K-Means gives the highest accuracy at a percentage of 65% and in which cobweb gives the least percentage of 5% which is not useful to implement.

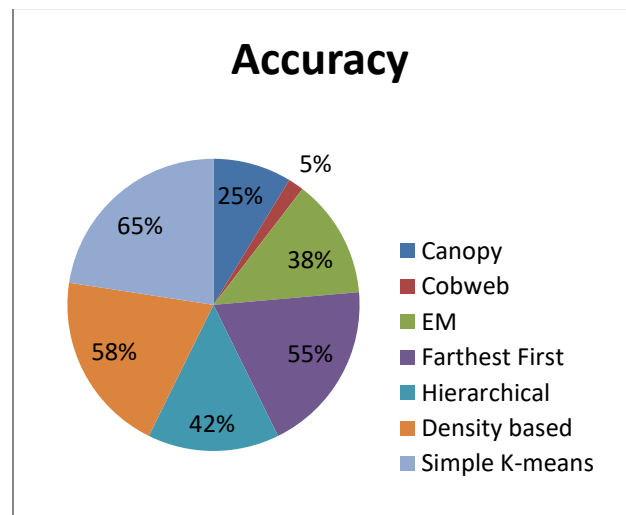


Fig. 9. Pie Chart – Accuracy

CONCLUSION

Allotting medicines to the patients who were tested with COVID-19 positive is presented. This is done by checking the specifications required in allotting the medicines such as their BP, diabetes, alcoholic and cancer details and finally proper medicine can be prescribed to the patients. Clustering algorithms are applied in which simple K-means gives an accuracy of 65% when compared to all other methods. This procedure presents an additional advantage of updating the application by adding other different types of diseases and allots medicines to patients updating the

information. This can be taken as reference for many applications.

REFERENCES

- [1] R. Punia, L. Kumar, M. Mujahid and R. Rohilla, "Computer Vision and Radiology for COVID-19 Detection," 2020 International Conference for Emerging Technology (INCET), Belgaum, India, 2020, pp. 1-5.
- [2] J. Laguarta, F. Hueto and B. Subirana, "COVID-19 Artificial Intelligence Diagnosis using only Cough Recordings," in IEEE Open Journal of Engineering in Medicine and Biology.
- [3] Swati Swayamsiddha, Chandana Mohanty, "Application of cognitive Internet of Medical Things for COVID-19 pandemic", *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, Volume 14, Issue 5, pp. 911-915, 2020.
- [4] de Souza W.M., Buss L.F., da Silva Candido D. "Epidemiological and clinical characteristics of the early phase of the COVID-19 epidemic in Brazil", medRxiv. Jan 2020.
- [5] Dong P., Ning Z., Obaidat M.S., "Edge computing-based healthcare systems: enabling decentralized health monitoring in Internet of medical Things", *IEEE Network*.30 Apr 2020.
- [6] [6] Vaishya R., Javaid M., Khan I.H., Haleem A., "Artificial intelligence (AI) applications for COVID-19 pandemic. *Diabetes & metabolic syndrome*", *Clin Res Rev*. 2020.