

Machine Learning Based Classification and Clustering Analysis of Efficiency of Exercise Against Covid-19 Infection

M. Sharmila Devi¹, Farooq Sunar Mahammad^{1*}, D. Bhavana², D. Sukanya³, TV. Sai Thanusha⁴, M. Chandrakala⁵ and P. Venkata Swathi⁶

¹Department of Computer Science, Santhiram Engineering College, India, E-mail: sharmila.cse@srecnandyal.edu.in

^{2, 3, 4, 5, 6}Department of Computer Science, Santhiram Engineering College, India

ABSTRACT

It is established that exercise will increase immunity against covid-19. This data can be statistically verified by employing a Support Vector Machine model and a K-Means method. The number of push-ups taken per day/week and the occurrence covid-19 infection can be modeled into a Support Vector Machine algorithm. Similarly the minutes of jogging per day and breathing difficulty occurrences can be modeled into a K-Means algorithm. The training data set will give a predictive model to assess the importance of exercise against covid-19.

Keywords: K-Means, Prediction, Corona, ML

1. INTRODUCTION

Corona virus is very dangerous for animals as well as humans. Humans are not known about this COVID-19 because it's a replacement genus before 2019. COVID-19 is spreading among people quickly, so it may become global epidemic problem. The main reason of this COVID-19 is Physical contact amid public. Various reasons for spreading this COVID-19 are hand contact, mucous contact, or breathe contact. It will mainly effect on public health and mainly economy also effects the daily life activities. To avoid spreading of its effect the main thing is to identify COVID Symptoms in the primary stage, so further we can control of spreading. Early Identification and analysis of COVID-19 reduce the load on Healthcare organizations. Prediction models that consist of several features to analyze the risk of infectivity are developed, within the hope of supporting medical staff around the world in triaging patients, especially within the condition of limited healthcare possessions. In The University of Virginia the researchers of Health organization put forward that daily exercise may drastically lessen the danger of acute respiratory distress syndrome, is the main reason for origin of demise in COVID-19 patients. In this paper, a machine learning model analyzes the efficiency of exercise with the references of classification and clusters the information. The risk of infection is analyzed with the machine learning model and predicts the Effectiveness of exercise during the infection. Clustering predicts the efficiency and reduces the death rates of infected Persons.

2. WORKING

The dataset consists of various exercises and their efficiency in combating with Covid-19 infection and other supporting data. The K-means clustering is employed here. The dataset is trained by finding number of clusters using the elbow method which helps us to find the 'K'. We employ the WSS i.e. Within Sum of

squares to find optimum clusters. WSS is the sum of squared distance between each cluster of groups and the centroid. This value is plotted and the Elbow point is found which gives us the number of optimum clusters. After measuring 'K' the centroid is established randomly and the distance between the points and centroid are measured and the centroid is repositioned. The process goes on until all the points are measured and the final centroid position is fixed and the grouping is done. Now we can compare this output get our required outputs.

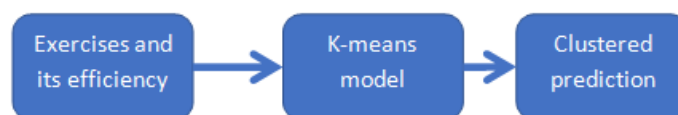


Figure 1: Block Diagram

2.1 K – Means

To solve clustering Problem one of the unsupervised learning algorithm which we will use in machine learning or data science is K-Means. Here, we are going to discuss about working procedure of an algorithm and k-means clustering implementation with Python. Without training of the data this algorithm will categorize unlabelled data based on the groups. Each cluster is connected with a centroid, in centroid-based algorithm. It reduces the distances in between the

data point and their related clusters. Unlabeled dataset can be given as input to the algorithm and divides that input into k-number of clusters and repeat the procedure till best clusters are not found in that. K value supposed to be determined by this algorithm. By iterative process Centroid or center points should be determined and allot each data point to the closet center point. By this Cluster can be created based on the nearest data points to the k-center.

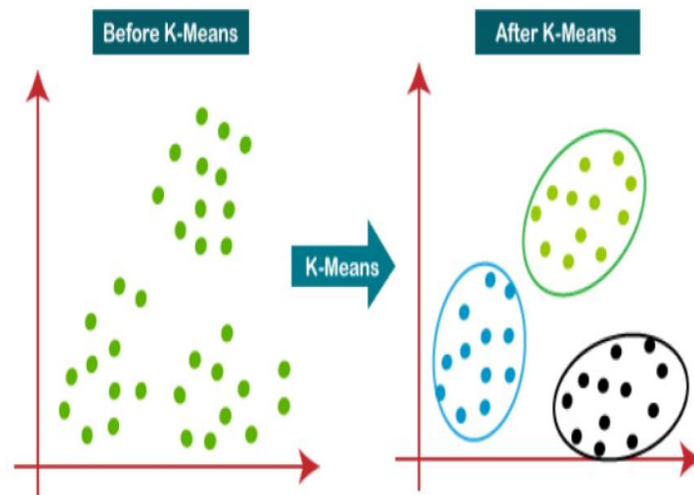


Figure 2. K-Means graph

Below steps will explain the working procedure of K-means Algorithm.

Step-1: Clusters can be indicated by the 'K'.

Step-2: Select indiscriminate K number of points or centroid. This can be supplementary from the input dataset.

Step-3: To the closest centroid each data point will be consigned, which will form the predefined K clusters.

Step-4: Determine the conflict and for each cluster put a novel centroid.

Step-5: Reiterate the third steps, which mean shuffle each data point to the new closest centroid of each cluster.

Step-6: If any relocation takes place, after that go to step-4 else go to conclude.

Step-7: The model is ready.”[4]

3. ARTIFICIAL INTELLIGENCE

The algorithmic flow presented in this paper falls under the field of Artificial Intelligence and Machine Learning. Presented below is an outline of those technologies.

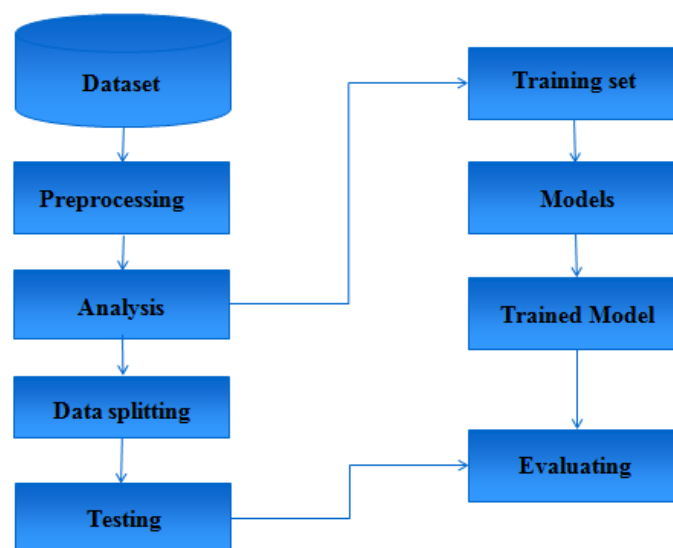


Figure 3. Flow Chart

Artificial intelligence is brain power established by machines, not like the natural intelligence displayed by human beings and animals. The distinction between the former and the last groups is often exposed by the ellipsis chosen. Strong AI is generally called as artificial general intelligence (AGI) and usual intelligence is called as artificial biological intelligence (ABI). We can say that AI is the study of Intelligent Agents and any device that acts on the environment and increase the chances to reach goals. Colloquially, artificial intelligence is frequently used to describe machines that mimic cognitive functions that humans associate with the human mind, such as learning and problem solving.

As Technology become progressively more proficient, tasks considered to require intelligence are often removed from the definition of AI, a phenomenon known as the AI effect. Whatever the work done till now is not about Artificial Intelligence. For example, optical character recognition is commonly expelled from things considered to be AI, became a regular technology. Modern machine capabilities generally classified as AI include successfully understanding human speech, challenging at the uppermost level in deliberate game systems as like chess and Go), and besides the games akin to poker, self-driving cars, intellectual direction-finding in content delivery networks, as well as military imitation.

Artificial intelligence was originated as an intellectual restraint in 1955, and in those days no type of encouragement from government and none other organization not sponsored any type of funds, so those days are called as AI Winter afterwards followed by new techniques and move forwarded with rehabilitated funding. A Professional chess player was defeated by Alpha Go in 2015, then another time AI came into Existence which pays attention all over the world. Many of the Researchers failed to communicate with the sub field of AI. These sub fields are based on technological deliberation. These sub-fields are based on technological deliberation, such as meticulous goals robotics or machine learning, the use of fastidious apparatus like logic or artificial neural networks, or profound idealistic differences. Sub-fields have also been based on communal characteristics.

The conventional tribulation of AI research comprises interpretation, acquaintance illustration, forecast, wisdom, natural language dispensation, discernment and the capability to move and maneuver objects. AGI is in the midst of the field's lasting aspiration and in this different type of methods embraces arithmetical methods, computational power and conventional representational AI. For geometric optimization, ANN and for different methods used in statistics, financial side several tools are used in AI by researchers. The AI field based upon computer science, information engineering, mathematics, psychology, linguistics, philosophy, and many other fields.

This Artificial Intelligence in simulated based on the individual intelligence and specifically described by the device. This elevates idealistic influences about the mind and the principles of creating artificial beings brilliant with intelligence like human. In the distant past these all have been indicated by myth, creative writing and thinking. Some of the researchers identified AI may become threat to humans if it is misused. On the Other side this AI will also increases the huge unemployment if the technology grows rapidly.

AI technique have practiced a renaissance subsequent concomitant advances in computer power, huge amounts of data, and speculative understanding in the twenty-first century and these Artificial Intelligence techniques plays vital role in many fields like Computer science, research, and product development.

4. MACHINE LEARNING

Machine Learning is mainly based on experience and it improves without human intervention by using the data, which may be labeled or unlabeled. We can say that Machine learning is the subfield of Artificial intelligence. A Model can be build based on some sample data well known as training or labeled data which is used to forecast or assessment devoid of being explicitly programmed in Machine Learning. These Algorithms mostly used in different areas like Healthcare, speech recognition, image processing, military, education, where the use of conventional algorithms will not give best results.

We cannot say that machine learning is statistical learning; this Machine Learning is intimately allied to computational statistics which highlights in taking forecast. The main application domains of this machine learning are mathematical methods and theory used for optimization. For unsupervised learning, tentative focus data analysis is done through data mining. Machine learning also called as predicative analysis based on the usage of business problems.

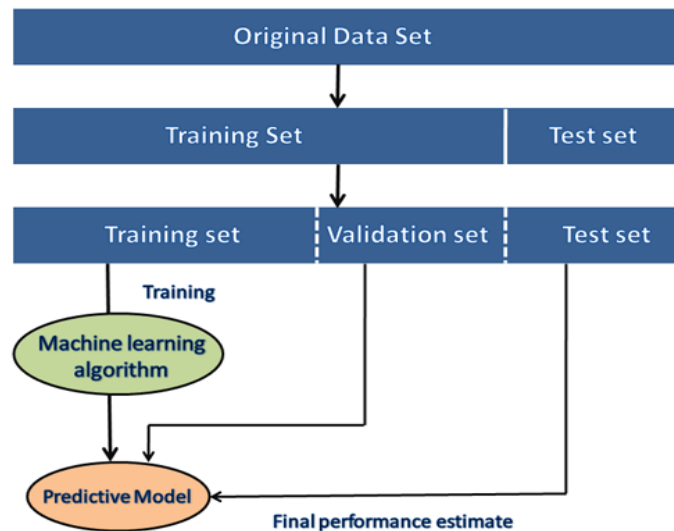


Figure 4. Machine learning Dataset

5. INDUSTRY 4.0

The Fourth Industrial Revolution uses of smart technology which is the enduring computerization of conventional development and engineering practices. For automation, enhanced communication and invention of smart machines which can evaluate and analyzing of all concerns without any involvement of humans intrusion possible by Internet of Things and Machine to Machine Communication.

Automation under Industry 4.0 has a particular schema or pattern at its outset. Presented below is how automation in the mass production industry as well as consumer level products is built in today's technological era.

The schema presented in Figure 5 has a lot of other components involved but the generic outline of it stands justifiable for all kinds of automation today.

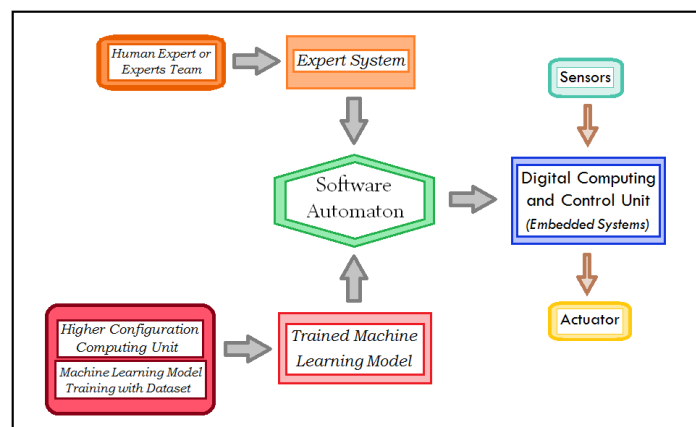


Figure 5. Schema of Automation

The software automaton of the conventional automation model, which is the status quo, was built by a human expert or a team of human experts till now. Human proficient's are not fully deliberated the software mechanization with the initiation of machine learning expertise. For the Software which was built by the human proficient's, the training in turn will be given in the outline dataset. By using the mathematical model the pattern amid the input and the output parameters of the dataset identified by machine learning software. The further electronic computing device uses this mathematical model as a operational software module.

This mathematical model is referred to as the 'trained machine learning module'. The software automaton of all the current digital embedded devices is a mathematical model that gives a numerical output for a numerical input based on arithmetic and logical conditions. This software automaton, as explained above can be either directly developed by a set of human experts by means of setting the boundary conditions themselves based on observation and requirement or can

be downloaded as an executable module from machine learning training systems that are trained with relevant dataset. In whatever way the software automaton is developed, it can be loaded onto the relevant embedded computing module that can be used for either sensor based closed loop automation or open loop automation.

The technological components of Industry 4.0 includes IoT, augmented reality, virtual reality, cloud computing, 3D printing, big data analytics, networking, data security, human-machine interaction and others. IoT is a very effective way to collect real world data. Sensors integrated with data acquisition and transmission systems can be placed anywhere and the collected data can be pre-processed if required and used as datasets to train machine learning models.

Cloud computing is employed for optimized utilization of computing resources. There are many third party vendors like Google and Amazon which are very reliable in terms of data security and speed of computation. These services offer companies and organizations a cheap and reliable way to harness the power of artificial intelligence and machine learning.

Big data analytics is the set of technological components involved with collecting, collating and managing large quantities of data for analytics and decision making. When so much data is involved, especially with third party service providers, data security plays an important role.

One of the paramount concerns about Industry 4.0 is the unemployment it can create due to powerful automations. The field of human-machine interactions and co-working has been a very developing field now to mitigate the above mentioned problem.

6. RESULTS AND DISCUSSION

Running/Push-ups	Breathing Disorder
1	20
3	17
4	15
5	14
8	12
10	11
12	10
14	9
17	7
18	6

Figure 6 Input Sample Dataset

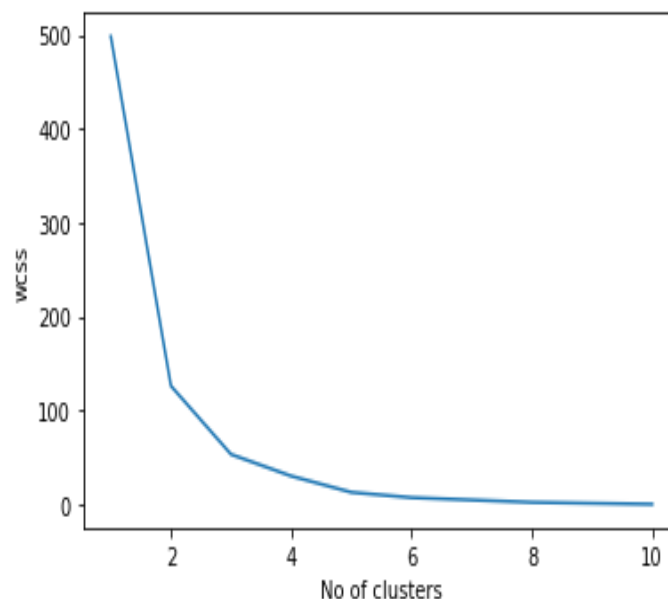


Figure 7. Output graph 1

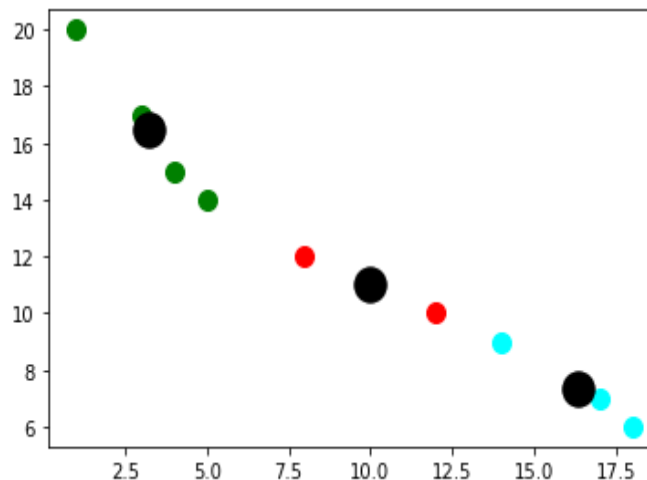


Figure 8. Output graph 2

This machine learning based predictive model is implemented employing Python programming language. The relevant library files have been included for execution of the code. The dataset was given as input and the predictive model obtained. The prediction model was tested and the results were satisfactory. The output graph of the machine learning model has been presented above. Support Vector Machine and K-Means algorithms have been implemented for this predictive model. Both the above classification and clustering models map the dependence of breathing disorders with exercise.

7. CONCLUSION AND FUTURE WORK

By this experiment, the unsupervised learning-based K-means clustering we are able to give out a clear cut interpretation and explanation from the predictive output. We can interpret that the advantage of various exercises helping us in combating the Covid-19 infection can be clearly stated out. This in turn explains the relation between exercise and ability to withstand Covid-19 infection. The future works can be carried out using an android application which can enable the user to track his overall progress of his/her exercise and its wellness can be measured.

REFERENCES

- [1] Nehal A. Mansour, Ahmed I. Saleh, Mahmoud Badawy, Hesham A. Ali. "Accurate detection of Covid-19 patients Based on Feature Correlated Naïve Bayes (FCNB) Classification strategy", Journal of Ambient Intelligence and Humanized Computing, 2021
- [2] Alazab M, Awajan A, Mesleh A, Abraham A et al (2020) COVID-19 prediction and detection using deep learning. Int J Comput Inf Syst Ind Manag Appl 12:168–181
- [3] Brinati D, Campagner A, Ferrari D, Locatelli M et al (2020) Detection of COVID-19 infection from routine blood exams with machine learning: a feasibility study. J Med Syst 44(135):1–12
- [4] Zhangmeng Xu, Yong Chen, Duoduo Yu and Dongdong Mao, "The effects of exercise on COVID-19 therapeutics: A protocol for systematic review and meta-analysis", Medicine: September 18, 2020 - Volume 99 - Issue 38 - p e22345.