

CORRELATION BETWEEN TEMPERATURE AND INCREASE IN COVID-19 CASES IN TELANGANA STATE

D.Lakshmi Padmaja¹, Medisetty Sujith², Sai Sruthi Bejagam³, Manish Reddy Morapally⁴
glpadmaja@gmail.com, medisujith@gmail.com, bejagamsaisruthi2000@gmail.com,
manishreddym5454@gmail.com

Abstract: The main aim of this paper is to know whether the temperature have any impact on the increase of corona virus. Covid-19, this name has brought a drastic change in our day-to-day life. People of Telangana have lived through 10 months of the Covid-19 pandemic and there might be more to come. Till date 2.8lakh official cases of covid-19 have been registered in Telangana and there may be many more which have gone unnoticed. In our busy life, we are neglecting our health and no one is maintaining a proper hygiene. And we have been more addicted to junk foods rather than nutrition, because of this reason covid-19 became a threat to our life. So proper precautions and awareness must be spread among people to avoid spread of virus. Fever, cough, breathing problems etc are the symptoms of this covid-19. If we neglect these symptoms it leads to a severe problem like pneumonia, kidney failure and eventually leads to death of that person. At this moment we don't have any vaccine to cure this disease, the only prevention or avoiding corona is to boost our immune system. To overcome this pandemic situation, firstly we need to know the important factors that increases in covid-19 cases. In this paper, machine learning techniques are used to identify how temperature varies with the increase of covid-19. Which means we find how the effect of temperature depends on the number of covid-19 cases in Telangana.

Keywords: Covid-19, K-means, Machine learning, Temperature.

I. INTRODUCTION:

Corona Virus is a family of related RNA viruses but only seven viruses can affect human. This covid-19 disease is due SARS-Cov-2 (severe acute respiratory syndrome corona virus 2), can easily spread from one person to another person. These viruses generally effect the respiratory track in humans and birds and they range from mild to lethal. Viruses generally replicate their RNA structure time to time and also depending on the conditions they are in. This results in the formation of new viruses. A person can transmit this SARS-Cov-2 virus within 48 hours before developing symptoms. Based on no symptoms and type of illness it stays 10-20 days in a human body. Covid-19 pandemic is the dark phase of the world, facing a global health crisis. The increasing death cases across the world is staggering, so we need to work together to decrease the spread of this virus. This pandemic not only effected the human health, it also gave an unexpected social-economic crisis.

Many of us think that temperature plays a vital role in development and transmission of the virus. Machine learning is best way to check whether this virus increases or decrease their growth based on climatic temperature. Here we use unsupervised learning, a machine learning technique which allow the model to understand on its own and concludes a particular pattern based on input factors.

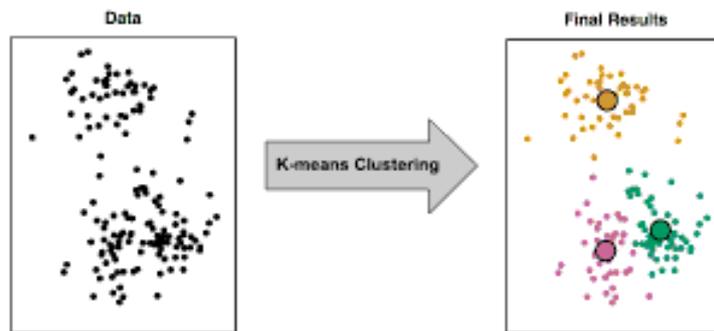


Fig. 1 K-means Clustering of the data

The steps that we follow to predict the output is as follows:

- 1: Collection of data
- 2: Data Pre-processing
- 3: Applying a classifier
- 4: Analysing the result

Brief explanation about the Introduction given in this section and, the remaining sections as follows. Literature survey, describes in Section II about survey that are already published in different journals. Section III explains about existing methods. In Section IV Challenges that are faced to identify the Covid 19 cases and its correlation. Section V explains in detail about proposed method, In Section V Results / Experimentation, in results, we presented our model resulted that are obtained after proposed method introduced. And in Conclusion section, we concluded about the paper based on the results.

II. LITERATURE SURVAY:

Amir Abdollahi and Maryam Rahbaralam published the Effect of temperature on the transmission of Covid-19: A machine learning case study in May 2020. In this research paper they used the Pearson's correlation coefficient and also employed a supervised machine learning techniques (Linear regression) to find the correlation between predictors (meteorological conditions) and target (no effected people). They found that the average temperature up to 6 previous days have a moderate negative impact with the daily increase in the infected people. According to Linear regression, every 1°C rise in average temperature there will be a decrease of about 200 infected people daily. They conclude by saying that they gave a clue to understand the transmission of Covid-19 and there are many other factors to define the transmission mechanism of Covid-19.

Zohair Malki, EI-Sayed Atlam, Aboul Ella Hassanien, Guesh Dagneu, Mostafa A, Elhosseini and Ibrahim Gad published a paper named "Association between weather data and Covid-19 pandemic predicting mortality rate: Machine learning approaches" on 2020 July 17. In this paper, they used many different supervised machine learning models and trained on the features like climate, census and health care. They concluded that the weather variables impact more in predicting the mortality rate when compared to the other variables.

Another paper name "A gradient boosting machine learning approach in modeling the impact of temperature and humidity on the transmission rate of Covid-19 in India" published by Lokesh Kumar Shrivastav, Sunil Kumar Jha on 04 November 2020. They used GBM (Gradient Boosting Machine) and concluded that the few atmospheric factors influence the increase in covid-19 cases especially Maharastra, Gujarath and Delhi.

III. EXISTING METHODS:

The various supervised machine learning algorithms like Logistic regression, Decision Tree, Support Vector Machine, Naive Bayes, k- Nearest Neighbours, and Random Forest can be used to find how temperature relates to increase in corona virus. And we can also use unsupervised machine learning like k-means algorithms, GBM. Deep learning, Artificial Intelligence can also help us to find the relation between temperature and increase in number of Covid-19cases.

IV. CHALLENGES:

This pandemic situation resulted in huge loss in every field. Adding to this vaccine is not yet found. It has become a challenging task for us to find the major cause that increasing the Covid-19cases in Telangana. With the technology that we have today, we can store huge amount of data. Based on the available data we can easily find out the main features using machine learning, deep learning etc. So that according to that results we can identify the main features and according to that if we take precautions, there will be a possibility to overcome this pandemic situation.

Virus may depend on temperature but we can't assure that temperature is the only factor for the increase of cases in Telangana. So, we used unsupervised machine learning techniques, in order to get a clear picture about this.

V. PROPOSED METHODS:**A. Data Collection:**

State wise confirmed cases and temperature data set of Telangana is taken. The State wise dataset consist of data, states, confirmed cases, recovered cases, reducedcases. Whereas temperature dataset consists of date, minimum and maximum temperature, collected from open data Telangana[5].

B. Data Pre-processing:

After collection of data and importing required packages we need to remove all unnecessary data, null values from the dataset and then we need to merge two different datasets that is number of Covid-19 case data set and temperature dataset of Telangana state based on the date attribute. This type of filtering and combining the data is called data pre-processing.

Finalized data set consist of date, state, confirmed cases, maximum and minimum temperature as attributes and mentioned in the Table 1.

Table 1: Finalized data set

	Date	State	Confirmed	Avg_temp	Temp_max	Temp_min
0	2020-03-02	Telangana	1	27.545161	33.458251	21.622071
1	2020-03-03	Telangana	1	27.360696	33.924788	20.796604
2	2020-03-04	Telangana	1	28.165789	34.745331	21.586248
....
210	2020-03-28	Telangana	187211	27.136163	31.157555	23.114771
211	2020-03-29	Telangana	189283	28.522835	33.021053	24.024618
212	2020-03-30	Telangana	191386	27.969355	32.422411	23.516299

213 rows x 6 columns

C. Applying Classifier:

Here, we use an unsupervised learning algorithm. Unsupervised learning is a machine learning technique which is used to study the model by itself and give a pattern as an output with the given input. It gives a pattern as an output by association and clustering the given inputs (From Fig. 1).

Clustering is a technique which groups the similar type of data points from the given input. We use k-means clustering as a classifier to develop our required model as shown in Table 2.

Apply K-Means on the above dataset

It is also known as iterative algorithm which divides the data points into k distant subgroups known to clusters.

1: Find we need to find the value of k

Here, we use elbow method to find the optimal number of clusters i.e k value.

2: Select k distinct data points randomly, we name that points as centroids.

3: calculate the sum of the squared distance between centroids and data points.

4: place each data point to the closest cluster and also calculate the centroids by taking the average of all data points which belong to each cluster.

5: Keep repeating 3,

4 steps until there is no change to the centroids of the clusters.

Table 2: Data set after clustering

	Date	State	Confirmed	Avg_temp	Temp_max	Temp_min	Cluster
0	2020-03-02	Telangana	1	27.545161	33.458251	21.622071	1
1	2020-03-03	Telangana	1	27.360696	33.924788	20.796604	1
2	2020-03-04	Telangana	1	28.165789	34.745331	21.586248	1
....
210	2020-03-28	Telangana	187211	27.136163	31.157555	23.114771	2
211	2020-03-29	Telangana	189283	28.522835	33.021053	24.024618	2
212	2020-03-30	Telangana	191386	27.969355	32.422411	23.516299	2

213 rows x 7 columns

K-means algorithm is applied for temperature and confirmed cases in Telangana, we got three clusters. All the number of confirmed cases in each cluster are aggregated.

Now we have selected the clusters with maximum and minimum corona cases.

From these clusters, we calculated the maximum and minimum temperature.

VI. Result / Experimentation:

The results that are obtained from the above model:

Data points before clustering is displayed in Fig. 2(X-Axis: Confirmed-Cases and Y-Axis: Temperature)

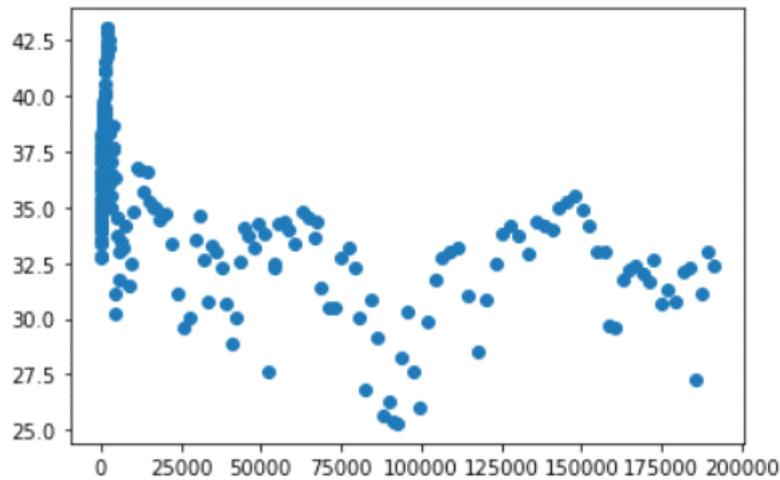


Fig. 2 Data points before clustering

Data points after clustering is displayed in Fig. 3(X-Axis: Confirmed-Cases and Y-Axis: Cluster-ID)

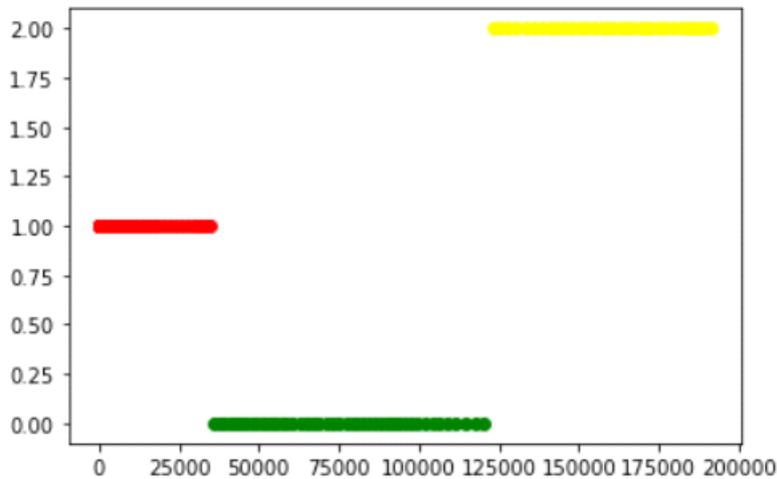


Fig. 3 Data points after clustering

From Fig. 3, we observe that more number of confirmed cases are there in cluster-id 2 and less number of confirmed cases are there in cluster-id 1

Table 3: Details of three cluster

Cluster-ID	Maximum Temperature (°C)	Minimum Temperature (°C)	Covid-19 Cases
0	34.8	25.3	3547434
1	43.0	29.6	594692
2	35.5	27.2	5082679

From the Table 3, we observe that a greater number of corona cases are in between the temperature 35.5 °C and 27.2 °C

Whereasa smaller number of corona cases are in between the temperature 43°C and 29.6°C.

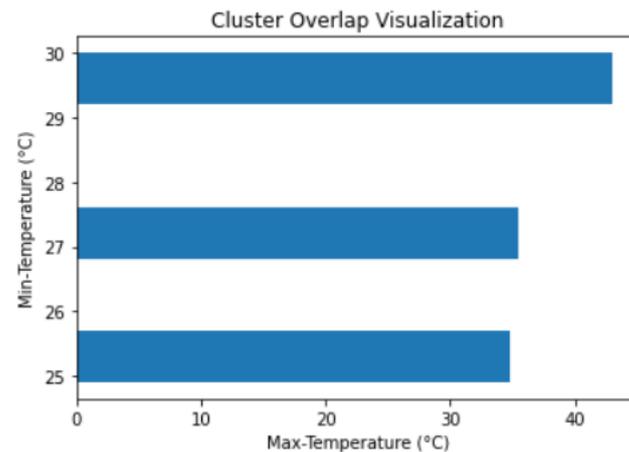


Fig. 4 Cluster overlap visualization

From Fig.4, the minimum and maximum temperature of the clusters are overlapping we cannot say that Temperatures is only the factor that influences the raise of corona cases.

VII. CONCLUSION:

From the above observations,we can definitely say that the range of temperature is overlapping, that means it is not a significant factor to find the reason for the increase of Covid-19cases in Telangana. Covid-19 cases may increase because of other factor like population density, immune power. Maintaining a proper social distance, boosting our immunity, wearing mask must be implemented in our life for not get effected to this Covid-19.

The machine learning model is built with the help of k-means clustering algorithm because it is easy to implement, can give efficient results for large data and have high-speed performance.

VIII. REFERENCES:

- [1] Mohammad khubeb sidduqui, Pradeep Gupta,Hafiz Iqbal, Ruben Morales-Menendez “Correlation between Temperature and COVID-19 (Suspected, Confirmed and Death) Cases based on Machine Learning Analysis” DOI: 10.22207/JPAM.14.SPL1.40, April 2020.
- [2] Amir Abdollahi, Maryam Rahbaralam “Effect of Temperature on the Transmission of COVID-19: A Machine Learning Case Study in Spain”DOI: 10.1101/2020.05.01.20087759
- [3] Zohair Malki, EI-Sayed Atlam, Aboul Ella Hassanien, Guesh Dagneu, Mostafa A, Elhosseini and Ibrahim Gad “Association between weather data and COVID-19 pandemic predicting mortality rate: Machine learning approaches ” DOI: 10.1016/j.chaos.2020.110137, 2020 July 17.
- [4] Milind Yadav,Murukessan Perumal and M Srinivas “Analysis on novel coronavirus (COVID-19) using machine learning methods” DOI: 10.1016/j.chaos.2020.110050, 2020 jun30
- [5] <https://data.telangana.gov.in/dataset/2020-telangana-weather-data>