COVID-19 Risk Assessment: Preparation and Response

1st Mayuresh B. Hooli
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

2nd Javad Birjandtalab Golkhatmi
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

3rd Alison Vicary
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

***Abstract*—We can help to curb the rate of the spread of this Virus by using the Ericson AI/ML knowhow. This identifies the counties that are hotspot for the virus. It identifies how quickly it can transfer to other counties. The short number of the testing kits in US is causing a huge problem. The government needs to prioritize the kit distribution. This approach helps to optimize the use of the kits. The problem is not addressed in the moment. We are using a disaster management approach which is the need of hour. We evaluate it with the actual value of cases in each county. We use Texas as the case study.**

***Index Terms*—COVID-19,Coronavirus, Risk Assessment, Spreading rate**

I. Introduction

A. Significance

COVID-19 has been declared a global pandemic. Thought the mortality rate is slightly lower than 5 percent, the rate of spread is something that hasn't been seen since the Spanish Flu in the early 20th century. The registered number of cases in USA is 174,684, but the numbers are based on the tests conducted. With a limited number of test kits available, we need to identify the counties that are hotspots for the spread of the virus. Once the hotspots are identified, we need to identify the rate of spread from the counties to the other counties. If these critical parameters can be identified, we can have some measures in place that will be crucial to fight this virus. (Reference need to be addedd)

B. Contribution

The COVID-19 Management is an intelligent system which identifies the critical areas (case study: Texas) where certain actions need to be taken. This system also ranks the counties based on the risk they pose to other counties for spreading the virus. This novel idea is based on implementing clustering alongside the PageRank algorithm. Clustering identifies 4 clusters based on the data set we prepared. PageRank algorithm uses the clustering prediction to identify which counties risk spreading the virus further. There is a shortage of medical personnel and testing kits, the categorization of counties helps in effective allocation of resources. The prediction of future affected counties will help in better preparedness to mitigate the virus. Ericsson is a pioneer in the field of communication. The communication technology developed by Ericsson would help understand the travel history of a county on an average to better prepare for the coming weeks

Identify applicable funding agency here. If none, delete this.

C. Related Works

...

II. Proposed Method

The idea is more useful for state and local government in terms of optimizing resources in this difficult time. Disaster Management tool which can get founding from the government. Government will get benefit from implementation to better controlling. People are interested to get a solution to see how safe different counties are. Ericson can show that its AI/ML know-how is valuable at these difficult moments.

A. Novelty

Everyone around the world is trying to figure out how to fight Corona-virus. The novel approach of using Clustering and Page Rank together would help us understand where the risk lies. Following the results given by this algorithm would help manage the fight against the pandemic.

B. Business Values

Ericsson's infrastructure would help predict the spread better by using the cell phone data to understand the travel history. This idea is useful for the Government which would help Ericsson get more Government contracts.

C. Page Rank Algorithm

. . .

D. Clustering

. . .

E. Data

We extract some meaningful features out of public census datasets and try to identify the hotspots. We will use data analytics and statistical analysis to study the effect of each indicator on spreading of the virus and predicting the spread into the hotspots. The census data that we use for hotspot identification is not used before.

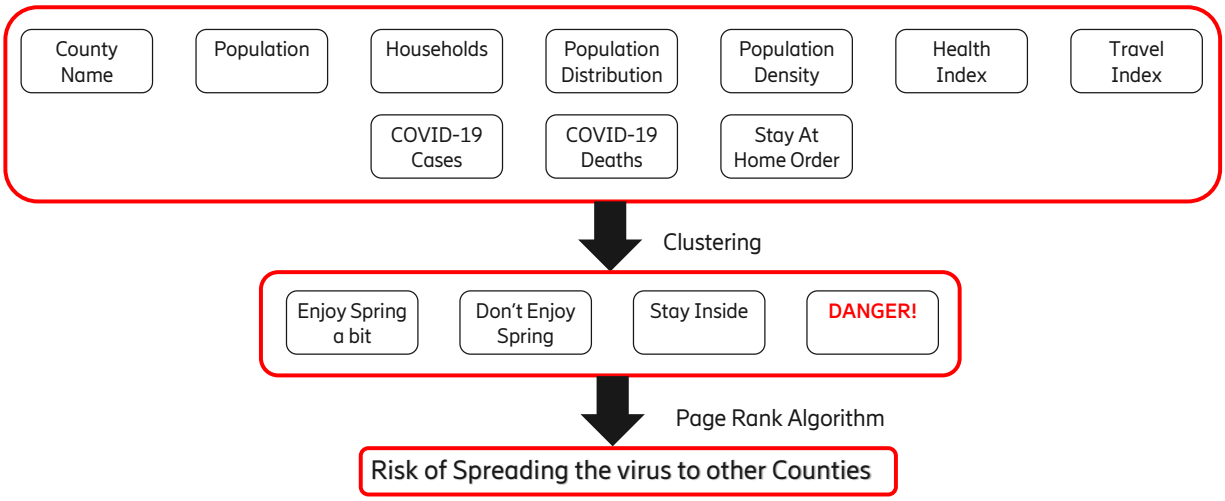

Fig. 1. The general view of our proposed methodology.

Here, we predict based on the counties. This can be done at a larger level for states and countries as well. But with more a more specified census data, we can predict better based on the counties. Multiple data sets are combined to get the data set required for this such as the census data, cleanliness data and the COVID-19 data.

The census data consists of the Geo ID of the county, FIPS, county name, the total male and female population as well as the population based on the percentages, males per 100 females and the same data distributed by age groups for every 5 years. We combine the categories to get a range of 0-15, 15-65 and 65 and above for the data set. We do not consider the gender ratio for the data set. We also consider the number of households from this dataset.

The next data set we use is the health index. The health index is based on a data set obtained from the healthiest communities' data set on US News. This data set contains information about the population health, equity, education, economy, housing, food nutrition, environment, public safety, community vitality and infrastructure of the county. We aggregate all this data to calculate the cleanliness index of the county to use in the data set. As we do not need the rest of the data now, we delete those columns and use only the aggregated column.

Next we obtain the population density from the land area. This is used because places there is lower social distancing and higher contact between people in places where the population density is higher. The population density is aggregated for the county rather than considering the concentration in a few places that is seen in various counties.

The next parameter considered is the travel index. The travel data is restricted hence we try different methods to find the travel history. We assume that every person with a passport has traveled. From this assumption, observe that there is a correlation between the travel data and the per capita income. There is also a correlation between the travel data and the education. Using these parameters, we calculate the travel index.

The next parameter we consider is the COVID-19 cases and deaths in each county. We also use the COVID-19 cases and deaths by population and household. Since these numbers are very small, we normalize them to get something we can use. We also use the COVID-19 stay at home labels. Since many of the Texas counties are rural counties with a small population density, people tend not to interact much with each other. We give the value 0 for counties that haven't implemented stay at home and 1 for counties that have implemented the stay at home.

F. Experimental Results

Using the parameters above, we use the k-means clustering algorithm. We get the counties based on 4 different clusters. The following diagrams give the results of the clusters in table I.

G. Spread Identification

Now we have identified the counties which are hotspots for the COVID-19. But there may be more counties that may be at risk of COVID-19 in the future. We base this on the closeness of the counties to each other. The counties that have a high concentration of COVID-19 cases may have an impact on other counties and raise the number of COVID-19 cases in that county. To identify the threat posed by other counties to the county in question, we use the PageRank algorithm. To use the PageRank algorithm, we use the matrix of neighbors. But as we see above, not all the neighbors are equal. Thus, we replace the value 1 with an addition of cases per unit area and deaths per unit area and then take a square root of them. Then we normalize this value to get the values in the matrix. We apply a PageRank algorithm to this matrix. Based on this matrix, we observe that the Dallas county has the highest risk of being infected even further while Wheeler County is at no risk.

TABLE I
EXPERIMENTAL RESULTS

Texas Counties	COVID-19 cases reported	Labels
Aransas	2	Enjoy Spring a bit
Archer	0	Enjoy Spring a bit
Austin	2	Enjoy Spring a bit
Bandera	0	Enjoy Spring a bit
Baylor	0	Don't Enjoy Spring
Elpaso	40	Don't Enjoy Spring
Hidalgo	28	Don't Enjoy Spring
Wichita	28	Don't Enjoy Spring
Nueces	22	Don't Enjoy Spring
Cameron	20	Stay Inside
Travis	200	Stay Inside
Denton	165	Stay Inside
Collin	134	Stay Inside
Fort Bend	119	Stay Inside
Galveston	70	Stay Inside
Bexar	157	Danger
Dallas	488	Danger
Harris	526	Danger
Tarrant	139	Danger

. .

III. FUTURE WORK

We plan to collect the cell phone data to better understand the travel history and aggregate based on county as a whole Extend this algorithm to the whole country by collecting data from all counties. We will use the GIS system to identify the zones on the map for better visualization. We will combine model with time series analytic. We will create a similar model for pandemic response and disaster response to help manage similar situations in the future.

IV. CONCLUSION

. . .

REFERENCES

[1] G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955.

[2] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[3] I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[4] K. Elissa, "Title of paper if known," unpublished.

[5] R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.

[6] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[7] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.