COVID-19 Spread prediction Based on Food Factors using Data Science

Y.Mahima
Informatics Institute of Technology
University of Westminster
Colombo,Sri Lanka
yasasymahima@gmail.com

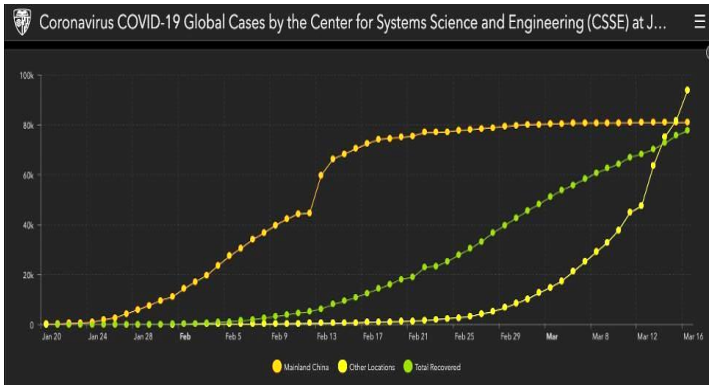
T.N.D.S. Ginige
Universal College Lanka
Monash University
Colombo,Sri Lanka
thepul@ucl.lk

***ABSTRACT* – Novel corona (COVID-19) disease is an infectious disease caused by a newly discovered coronavirus. Covid-19 was started from Wuhan China and present it is spread all over the world. Presently the USA is affected by the Covid-19 virus hardly. Present there are more than 4.0 million confirmed cases and more than 200 thousand happens because of the Covid-19 virus.** [1]**. However there is no medicine found by the researchers yet for the virus. According to the WHO (World Health Organization), the COVID-19 virus is spreading by several methods. And for protection from the COVID-19 virus their suggestion is community distance. Moreover doctors and other medical instructors advise concentrating more on doing sanitizations often. Presently there are lots of researches doing for identifying reasons, make a medicine in this area. When it concentrates on the spread of the COVID-19 virus it cloud be able to identify that there is a difference between Asian and European countries. As an example from Asian countries only china is affected by the COVID-19 virus majorly. However European countries like Italy, Spain has the highest death rates more than china. Mainly south Asian countries have a low confirm rate of the COVID-19 virus.** [2] **For this the main reason what here in this research identified is food used by each region. We know that people in European countries eat fast foods and they have used alcohol more than the people in Asian countries.** [3]

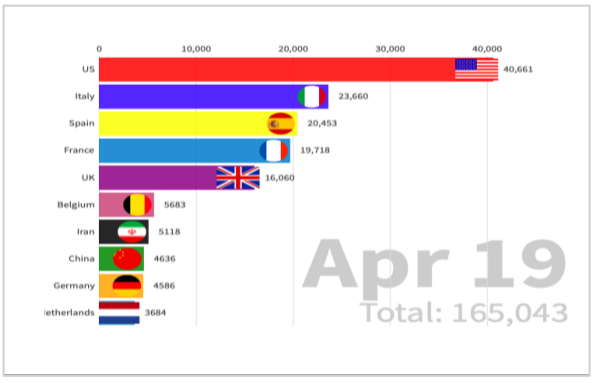
Therefore this research was done to identify a pattern between the spread of the COVID-19 virus and the Food factors used by the people. So here in this research using data science. Here it does some data modeling part to visualize a pattern between the spread of the COVID-19 virus and food factors. Thereafter using a machine learning model here it implement a prediction model. Through this research, it could be able to understand that there is a clear pattern between the spread of the virus and the food factors. Moreover this research will be a novel approach to evaluate the spread of the COVID-19 in selected areas.

I. INTRODUCTION

World Health Organization has declared the Novel Corona Virus disease in 2019 a pandemic. This was started on December 31 by a Cluster of patients found in the city of Wuhan and Hubei province in China. The Impact of the Coronavirus is bigger than the previous outbreaks like SARS and MERS. As of 15 May 2020, there are over 4 million cases have been identified and over 200 thousand deaths were reported globally in 188 countries. Moreover, nearly 1.5 million were recovered. By the below graph it could be able to clearly understand the spread of the virus from the beginning. [4] [5]

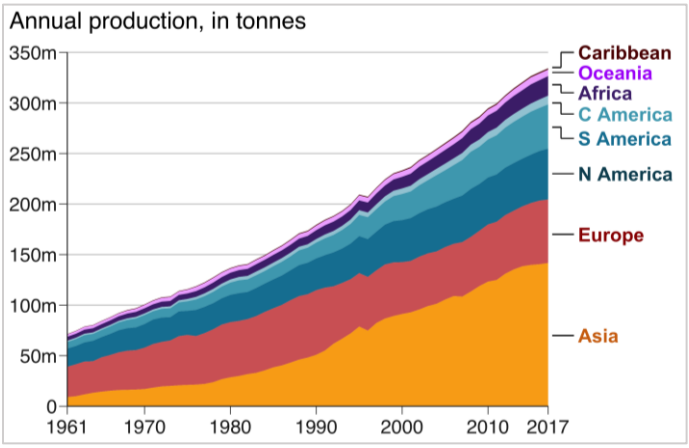


When it concentrates on the spread of the death rates of the COVID-19 virus there can understand that the spread of the virus is faster in European countries and the USA when comparing to the Asian countries. Coronavirus was started from china. But there were only nearly 82 thousand confirmed cases and 4.5 thousand deaths reported in China. [6] But in Italy, Spain Britain the confirmed cases and deaths are several times greater than China. By below bar char it can clearly understand that most of the European and Western countries affected by the virus in a very fast manner. [7]



For the difference between the increment of the COVID-19 cases in western and Asian countries there are several reasons identified by the previous researchers. The main thing they identified in lack of social distance in the daily activates of the people. The next thing is the weather. Most of Europe and western countries have cold weather. But Asian countries have hot weather. Hot weather is able to kill the COVID-19 virus in the environment.

However from this research it will concentrate on the Food Factors which causes for the spread of the COVID-19 virus. When it compare the foods taken by the Asian and Western people there is a difference. Western people eat fast foods and foods include meats and using alcohol more when comparing to the Asian people. The below graph shows how meat product usage grows in each region according to the year. [3]



From those factors, there can assume that there is a probability that food factors also cause for the spread of the virus. Therefore from this research it is able to recognize a pattern between food factors and the spread of the coronavirus.

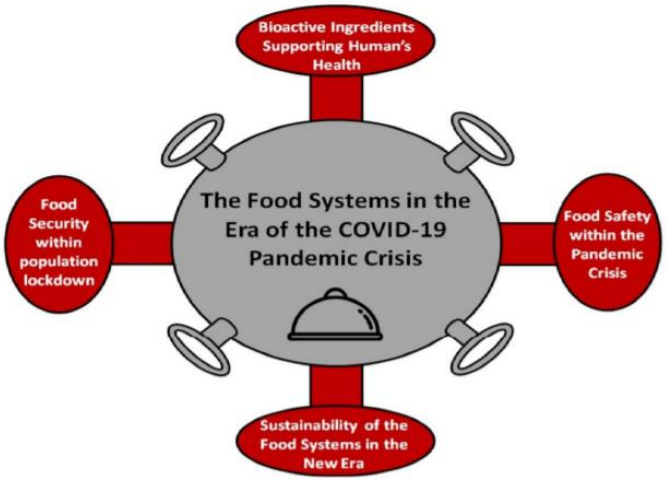
As mentioned about to predict how the COVID-19 virus spread with the Food factors here this research used a data science component. First to identify how the confirmed cases change according to the food factors in each country here this research used data modeling concepts. Then to identify a pattern and predict the spread of the virus according to food factors usage here this research used some algorithmics approaches and trained a model.

So this is a brief summarization of this research. Here this research concentrate on a less concentrated area on the COVID-19 virus previously. Finally in the implementation process of this research here it got a 63% accurate data science model to predict COVID-19 spread prediction relevant to the usage of food factors. This research will take a novel approach to analyze the spread of the COVID-19 virus and it helps to self analyze and save from the COVID-19 outbreak. In the next chapter, it will discuss several competitive pieces of research done in this area.

II. LITERATURE REVIEW

Novel coronavirus became a world problem now. Therefore there are lots of research done in this area. There are researches to predict if a person has a virus or not, methods that virus spread and predict the spread of the virus..etc. As mentioned in the introduction chapter here it analyzes how the COVID-19 virus spread according to the usage of food factors. Moreover to analyze here it used a data science approach. When it concentrates on previous researches done in areas like nutrition, foods related to the spread of the COVID-19 Virus and reasons for the spread of the virus several types of research can find.

There is a research by Charis M. Galanakis named “The Food Systems in the Era of the Coronavirus Pandemic Crisis”. Here in this research he had concentrate on the bioactive ingredients of foods and herbs for the support of the human immune system against infections before discussing the possibility of COVID-19 transmission. Here this research had clearly discussed how the bioactive ingredients in the food support for the Human Immune system to protect from the coronavirus and further it had discussed sustainable food systems for the virus. Other than these topics this research focused on food security within the population lockdown which is not relevant to this research.



Here in this research it discussed how nutrients like Vitamin C, B-Carotine helps to the Immune system of the Body. Moreover it discussed how the herbs and Chinese medicines support for Immune system by using pieces of evidence like usage of those herbs in other outbreaks like SARS. [8]. The difference between this research and when comparing to our approach here it does not predict the spread of the COVID-19 virus with the food factors. Moreover here in this previously mentioned research it is not a data science approach.

There is another research done related to foods and the Covid-19 virus by two researchers named Farah Naja, Rena Hamadeh to analyze the food security in the world because of the COVID-19 virus. Here they were concentrated on how food security can be improved by nationally and globally. [9]. There is another research done by HLPE (High-Level Panel of Experts on Food Security and nutrition) Italy which is almost similar to the previous one. Here in this research, it will also concentre on Food security in Italy and the increment of the food demand. [10]

However both of these researches do not concentrate on how the food factors, food usage causes the spread of the COVID-19 Virus. But the area here in this research concentred by the author is how the distribution of the COVID-19 virus change by the usage of the food factors. So these are the researches done related to foods and the COVID-19 virus.

There are several pieces of research done to predict the distribution of the COVID-19 virus and causes for the virus. There is research done by several researchers to predict the spread of the COVID-19 virus using machine learning named "Preparation analysis and prediction of the COVID-19". Here they were used time-series analysis to predict the spread of the COVID-19 virus. Moreover here they got a good accuracy rate for the model. [11].However the difference between this and the proposed research by the author is here they used time series analysis but in the authors' research it will analyze the food factors which cause the COVID-19 and predict the spread of the virus using that. Rather than predicting the spread here this research will concentrate to identify how the food usage of the people causes the spread of the virus.

There is another research done to predict the COVID-19 distribution in African countries. Here they used the SIER and MH-algorithm based model for the prediction. Moreover here in this research Epidemic controlling measures including risk level classification were proposed. When comparing to the authors' research here this research also analyzes the past data and does a time series analysis for the prediction. [12] . There is another research done by an Indian researcher Arti .M.K to modeling and the prediction of the COVID-19 spread in India. Here in this research, it used a Tree-based approach to predict the spread of the virus. Here in this research it had also analyzed how the lock-down and isolation techniques cause to minimize the spread of the virus. [13].

So these are some related researches done to predict the spread of the COVID-19 virus and how food factors related to the spread of the virus. From these previous research approaches it can understand that there are very few researches done to analyze how the people's food usage causes the Covid-19 virus and what types of foods which support the humans Immune system to secure from the COVID-19 virus. Therefore this research takes a novel approach to analyze and predict how the spread of the COVID-19 virus change according to the food factors. In the next chapter it will clearly describe the proposed method and materials of the research.

III. METHODS AND MATERIALS

Here in this research it proposed a system that analyzes the spread of the COVID-19 virus based on the food factors used by the people and predicts the spread of the virus by that. For the implementation as the main technique here it is going to use a data science. Since there is an analysis section the implementation process of the research can divide into two main parts. They are

1. Data Modeling Process
2. Covid-19 Spread Prediction Model

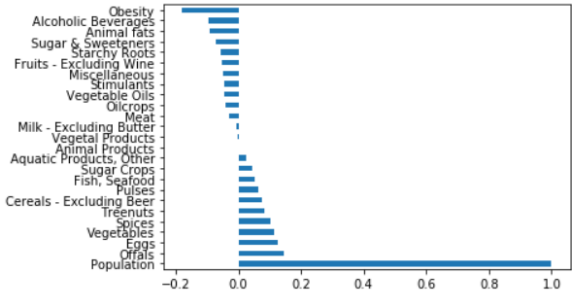
Therefore the implementation process will discuss according to the above sections briefly.

For both data modeling and for the prediction here it used a country-wise data set which includes a COVID-19 virus spread and the use percentage of each food factor according to the population. BY using this here first analyze how the spread of the virus happens according to each food factor. For the implementation of this the programming language used in Python. Therefore to the data modeling here it used several libraries and packages like matplotlib , plotty ..etc. The results which get from the data modeling section will discuss in the next chapter

The most important part of the research is predicting the spread of the COVID-19 virus according to the food factors. For that here also used the same data set. The implementation process of the prediction has two main sections. They are

1. Data Pre-processing
2. Choosing the correct model and do the prediction.

In the preprocessing here it remove null values and encode the strings. Moreover for improving the accuracy of the model here it standardized the data. Moreover here it analyze the correlations



By the above image it can understand that the population is the most correlated attribute. Therefore here it remove the highest correlated variable (Population) from the data frame because the model mostly depends on that attribute.

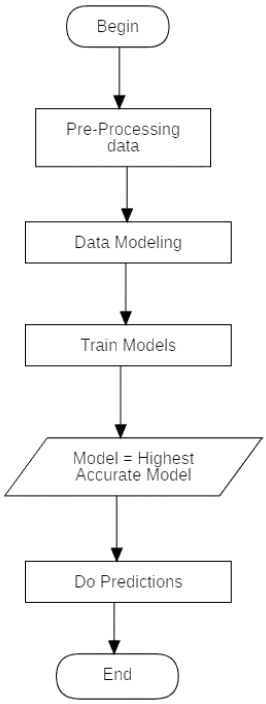
In choosing the correct algorithmic approach process here it used the regression model approach because here it needs to predict a specific value. Therefore here it analyzes several regressions models from the Scikit learn libraries and choose the

```
algorithms = []
algorithms.append(('LinearRegression', LinearRegression()))SSS
algorithms.append(('BaggingRegressor', BaggingRegressor()))
algorithms.append(('RandomForest', RandomForestRegressor()))
algorithms.append(('KNeighbours', KNeighborsRegressor()))
algorithms.append(('Boosting',XGBClassifier()))
```

best accurate model for the prediction.

The above image shows the used algorithms to choose the best accurate model. So here the BaggingRegresseor model gives the highest accuracy. Therefore the predictions here this research use the BaggingRegresssor model. The results of each model will discuss in the next chapter further.

This is the implementation process of the research to analyze and predict the impact of food factors to the distribution of the COVID-19 virus and predict the spread based on that, The whole process of the implementation of the research could be clearly described by the below flow diagram briefly.



Therefore this is the main process of the research implementation. In the next chapter it will clearly describe the chosen data set, the result of the data modeling part, and the accuracy of the trained models for the spread prediction based on food factors used.

IV. DATA AND RESULT

References

[1] WorlOmeter, "COVID-19 CORONAVIRUS PANDEMIC," WorlOmeter, 20 5 2020. [Online]. Available: https://www.worldometers.info/coronavirus/. [Accessed 20 5 2020].

[2] BBC, "Coronavirus: Five of the countries most at risk from famine in 2020," 22 4 2020. [Online]. Available: https://www.bbc.com/news/world-52379956. [Accessed 16 5 2020].

[3] BBC, "Which countries eat the most meat?," 4 2 2019. [Online]. Available: https://www.bbc.com/news/health-47057341. [Accessed 16 5 2019].

[4] Physiopedia, "Coronavirus Disease (COVID-19)," Physiopedia, 2020. [Online]. Available: https://physiopedia.com/Coronavirus_Disease_(COVID-19). [Accessed 16 5 2020].

[5] T. Nace, "Coronavirus Map: How To Track Coronavirus Spread Across The Globe," 16 3 2020. [Online]. Available: https://www.forbes.com/sites/trevornace/2020/03/16/coronavirus-map-how-to-track-coronavirus-spread-across-the-globe/#34308763f980. [Accessed 16 5 2020].

[6] Worldometer, "Coronavirus Cases China," Worldometer, 16 5 2020. [Online]. Available: https://www.worldometers.info/coronavirus/country/china/. [Accessed 16 5 2020].

[7] B. J. page, "The covid-19 pandemic in two animated charts," 27 3 2020. [Online]. Available: https://www.technologyreview.com/2020/03/27/950263/the-covid-19-pandemic-in-two-animated-charts/. [Accessed 16 5 2020].

[8] C. M. Galanakis, "The Food Systems in the Era of the Coronavirus," MDPI, Switzerland, 2020.

[9] R. H. Farah Naja, "Nutrition amid the COVID-19 pandemic: a multi-level framework," Springer, 2020.

[10] HLPE, "Impact of COVID-19 on Food Security and Nutrition," Rome, 2020.

[11] C. Z. PengaYiming, "Propagation analysis and prediction of the COVID-19," *Infectious Disease Modelling,* vol. 5, pp. 282-292, 2020.

[12] A. l. o. o. panelZebinZhao, "Prediction of the COVID-19 spread in African countries and implications for prevention and control: A case study in South Africa, Egypt, Algeria, Nigeria, Senegal and Kenya," *Science of The Total Environment,* vol. 729, 2020.

[13] A. M.K., "Modeling and Predictions for COVID 19 Spread in India," India, 2020.