

IDENTIFYING COVID-19 WITH NORMAL SICKNESS USING LOGISTIC REGRESSION

Dr. M. RAJESWARI¹,
Assistant Professor,
Department of B.Com(A&F and BA),
PSGR Krishnammal College for Women, Coimbatore, India.
rajeshwarim@psgrkcw.ac.in

S. VARSHASRI²
UG Scholar,
Department of B.Com(Business Analytics),
PSGR Krishnammal College for Women, Coimbatore, India.
varshasrisrikanth@gmail.com

ABSTRACT

Coronavirus causes symptoms like fever, body ache, dry cough, fatigue, chills, headache, loss of appetite, and loss of smell. In some people, COVID-19 causes severe symptoms like high fever, severe cough, and shortness of breath, which often indicates pneumonia. The Machine Learning (ML) model must be made to examine the information and it accepts that once contaminated individuals are not immune to the illness everlastingly, thus COVID infection happens once and afterward it can return again whenever with cutting edge changes in its qualities.

Keywords: Logistic Regression, Other diseases, Covid-19, prediction.

I. INTRODUCTION

Logistic regression may be a statistical analysis method to predict a knowledge value supported prior observations of a dataset. A data set can be a collection related data which can be accessed individually or in combined manner . The logistic regression is to estimate the number of possibilities of events, including determining a relationship between features and thus the probabilities of particular outcomes.

II. OBJECTIVE

The main objective is to predict whether the people are affected by covid-19 who were already affected by other diseases. People with other diseases like sugar, cholesterol, asthma, sinus will be easily affected by the coronavirus as they are already weak and have less immune system. This objective is done through supervised learning. The supervised learning has many algorithms from which logistic regression is used to predict the objective.

III. RELATED WORK

The first pneumonia case of unknown cause was found on the brink of a seafood market in Wuhan, the capital city of Hubei province, China, on Immaculate Conception , 2019. In late December 2019, patients with similar pneumonia were reported . The pneumonia was later identified to be caused by a new coronavirus (severe acute respiratory syndrome coronavirus 2, or SARS-CoV-2), later named Coronavirus Disease 2019 (COVID-19) by the World Health Organization (WHO).[6] China declared a Level 1 emergency response, the highest level public health response, to the COVID-19 outbreak on January 15 2020, causing the implementation of control measures nationwide.[8]

From early March, 2020, population-level non-pharmaceutical interventions (NPIs) to scale back SARS-CoV-2 transmission were introduced in many countries suffering from COVID-19, and these have included school closures, bans on public events, restrictions on gathering sizes, and requirements to remain in reception . Since the start of May, 2020, several countries have begun to lift a number of these NPIs, and a few countries have witnessed a second surge within the number of reported COVID-19 cases.[4] The novel coronavirus is leaving a deep socio-economic impact globally. Due to the convenience of virus transmission, primarily through droplets of saliva or discharge from the nose when an infected person coughs or sneezes, countries which are densely populated need to be on a higher alert.[1]

Computed tomography (CT) is a test that provides a window into pathophysiology that could shed light on several stages of disease detection and evolution [2] [10]. Generally, testing to find covid19 positive cases relies heavily on Reverse Transcription-Polymerase Chain Reaction (RTPCR), which is time consuming and has false-negative error. Thus, developing new approaches for detecting patients at a faster rate with higher accuracy is a matter of importance [5]. Even though CT and RT-PCR are most often concordant, CT can also detect early COVID-19 in patients with a negative RT-PCR test[9].

It is spreading through the bodily droplets in the air. Common symptoms include fever, tiredness, and dry cough. Along with these symptoms, a patient also experiences shortness of breath, aches and pains and sore throat. Very few people have experienced diarrhea, nausea or a runny nose. People having high fever, cough or difficulty in breathing should call their doctor and seek medical help immediately. Human to human transmission is exponentially increasing the counts of the infected people. The incubation period of this disease is 1–14 days or even longer [7].

First in the category are the elderly, who are highly susceptible to the virus. Statistics show that because of the weak immune system the elderly succumb to the disease easily. The second category is that of the children. As the immune systems of young children are still under development, the children are at higher risk. The third category is that of the people who have diseases like diabetes, high BP, asthma, cancer, cardiovascular disease, etc. As their immune systems have been compromised already due to a prevailing medical condition, these people become easy targets. Infections experienced by the third category of people can be fatal [3].

IV. METHODOLOGY

A. DATA MINING

Data mining is a process to convert raw data into useful information which is used by the companies. By using this software, companies can learn more about their customers to develop more effective marketing strategies, increase sales and decrease costs. Data Mining has different methods, one of them is classification. Under classification, Logistic regression is used to predict the COVID dataset.

B. LOGISTIC CLASSIFIER

Logistic classifier is a method of supervised machine learning. The logistic regression is a machine learning algorithm and it is used to describe the data and to explain the relationship between one dependent binary variable and one or more ratio-level independent variables. Logistic regression uses an equation which is extremely very similar to rectilinear regression. Input values are combined linear which use coefficient values to predict an output value. The algorithm is composed of following steps,

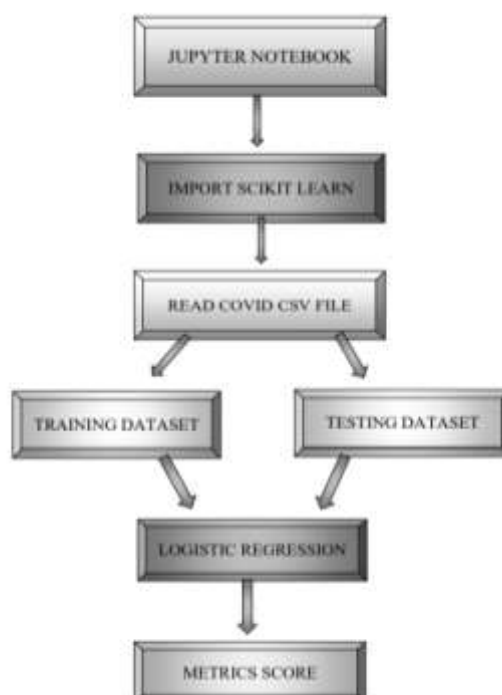
Step 1: Get a dataset and read it.

Step 2: Split the dataset into training and testing dataset.

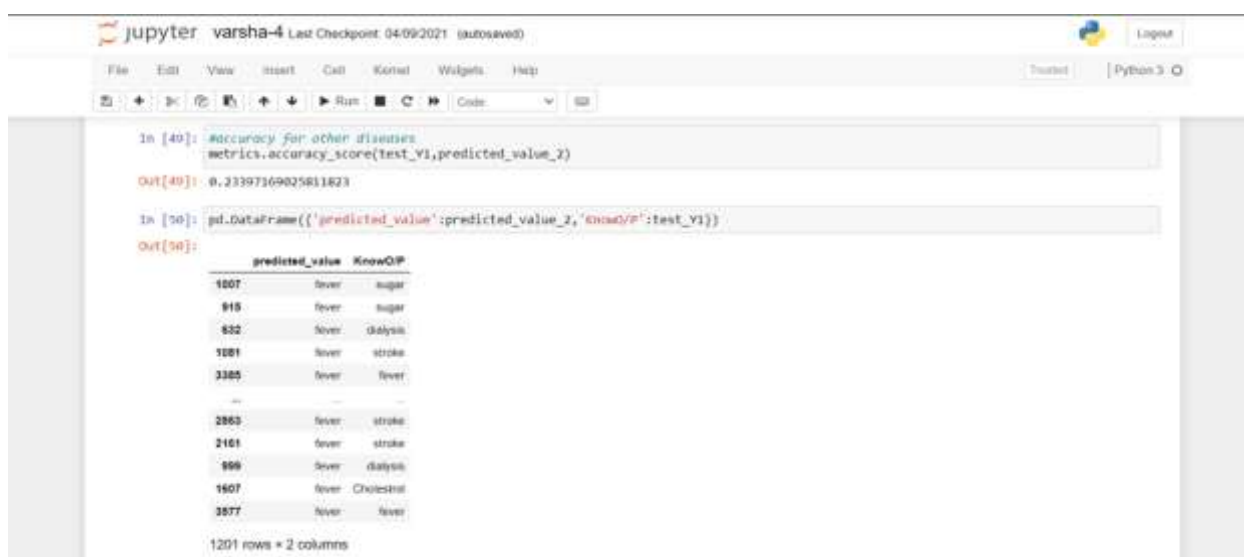
Step 3: Calculate the accuracy score to get the ratio between the number of correct predictions and all predictions made.

Step 4: Make a prediction using the classifier logistic regression.

C. FLOW DIAGRAM



V. RESULT



```

jupyter varsha-4 Last Checkpoint: 04/09/2021 (autosaved)
File Edit View Insert Cell Kernel Widgets Help
Python 3

In [49]: #accuracy for other diseases
         metrics.accuracy_score(test_Y1,predicted_value_2)
Out[49]: 0.23397369025811823

In [50]: pd.DataFrame({'predicted_value':predicted_value_2,'KnowOP':test_Y1})
Out[50]:

```

	predicted_value	KnowOP
1007	fever	sugar
915	fever	sugar
832	fever	diabetes
1081	fever	stroke
3385	fever	fever
...
2863	fever	stroke
2161	fever	stroke
999	fever	diabetes
1607	fever	Cholesterol
3577	fever	fever

1291 rows × 2 columns

The Most common symptoms of covid-19 are fever, dry cough, tiredness, the less common symptoms are aches and pains, sore throat, diarrhoea, conjunctivitis, headache, loss of taste or smell, a rash on skin, or discolouration of fingers or toes, the serious symptoms are difficulty breathing or shortness of breath, chest pain or pressure, loss of speech or movement. The prediction of covid-19 with the variable Other diseases is the result found in the above figure. For example, The coronavirus enters the cell only in the presence of cholesterol and GM1 in the form of lipid rafts, in culture. So the person who has cholesterol will have a severe condition. Therefore, the persons who have other diseases are at high risk because they may easily be affected by coronavirus. The metrics score for the other diseases is 23% in approx., which shows the prediction accuracy level. Mostly the persons who are under normal sickness have high chances of getting affected by coronavirus. Predictions made using this regression does not necessarily involve predicting the future but it predicts the mean of the dependent variable given specific values of the dependent variables. That is, one independent variable to predict the dependent variable. Here both of these variables are measured at the same time.

VI. CONCLUSION

In this paper, Jupyter Notebook is the tool which was used to predict if the people may or may not be affected by covid-19 when they are already affected by other diseases like common cold, cholesterol, stroke, sugar etc. using logistic regression. The medical field students can use this technique to get more knowledge about the people who are at high risk. This prediction helps the doctors as well as the public to be aware of the coronaviruses which were spreading all over the world.

REFERENCE:

- Bai Y., Yao L., Wei T., Tian F., Jin D.-Y., Chen L., Wang M. Presumed asymptomatic carrier transmission of COVID-19. JAMA. 2020 [PMC free article] [PubMed] [Google Scholar] [Ref list]
- Fang, Y. et al. Sensitivity of chest CT for COVID-19: comparison to RT-PCR. Radiology. <https://doi.org/10.1148/radiol.2020200432> (2020).
- Fong SJ, Li G, Dey N, Crespo RG Herrera-Viedma E. Finding an accurate early forecasting model from small dataset: a case of 2019-ncov novel coronavirus outbreak. arXiv preprint arXiv:2003.10776. 2020
- Li, Y. et al. Te temporal association of introducing and lifting non-pharmaceutical interventions with the time-varying repro-duction number R of SARS-CoV-2: a modelling study across 131 countries. Te Lancet Infectious Diseases 1–10, [https://doi.org/10.1016/S1473-3099\(20\)30785-4](https://doi.org/10.1016/S1473-3099(20)30785-4) (2020).
- Liang T. Handbook of covid-19 prevention and treatment. The First Affiliated Hospital, Zhejiang University School of Medicine Compiled According to Clinical Experience. 2020;68
- Qiu, Y., Chen, X. & Shi, W. Impacts of social and economic factors on the transmission of coronavirus disease 2019 (covid-19) in China. J. Popul. Econ. 33, 1127–1172. <https://doi.org/10.1007/s00148-020-00778-2> (2020).

7. World Health Organisation online available on

<https://www.who.int/emergencies/diseases/novel-coronavirus-2019/>

8. Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *Lancet* 2020-[https://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(20\)30260-9/fulltext](https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(20)30260-9/fulltext)

9. Xie, X. et al. Chest CT for typical 2019-nCoV pneumonia: relationship to negative RT-PCR testing. *Radiology*. <https://doi.org/10.1148/radiol.2020200343> (2020).

10. Yu, M. et al. Thin-section chest CT imaging of coronavirus disease 2019 pneumonia: comparison between patients with mild and severe disease. *Radiol. Cardiothorac. Imaging* 2 <https://doi.org/10.1148/ryct.2020200126> (2020).

