**ANALYSIS OF COVID-19 BASED UPON SYMPTOMS**

***A software requirement specification report submitted in partial fulfillment of the requirements***

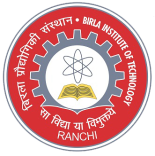
***for the award of the degree of***

**MASTER OF COMPUTER APPLICATIONS**

**(CA 656)**

BY

**Meghaj Kumar Mallick (MCA/25017/18)**



**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING**

**BIRLA INSTITUTE OF TECHNOLOGY, MESRA**

**JAIPUR CAMPUS, JAIPUR**

**SP-2021**

**INDEX PAGE**

|  |  |  |
| --- | --- | --- |
| **SR NO.** | **TOPIC** | **PAGE NO.** |
| 1 | PROJECT DEFINITION | 3 |
| 2 | INTRODUCTION | 3-4 |
| 3 | PROBLEM STATEMENT | 4 |
| 4 | AIM | 4 |
| 5 | OBJECTIVE | 4 |
| 6 | SCOPE & FUTURE WORK | 5 |
| 7 | SOFTWARE DESIGN | 5 |
| 8 | DATA FLOW DIAGRAM | 6 |
| 8 (a) | Context diagram(DFD Level 0) | 6 |
| 8 (b) | Context diagram(DFD Level 1) |  |
| 9 | USE CASE DIAGRAM |  |
| 10 | CLASS DIAGRAM |  |
| 11 | NON FUNCTIONAL REQUIREMENT |  |
| 12 | LIMITATIONS |  |
| 13 | CONCLUSION |  |

1. **PROJECT DEFINITION**

The recent outbreak of the respiratory ailment COVID-19 caused by novel Corona Virus SARS-Cov2 is a severe and urgent global concern. In the absence of effective treatments, the main containment strategy is to reduce the contagion by the isolation of infected individuals; however, isolation of unaffected individuals is highly undesirable.

We developed a model that employed supervised machine learning algorithms to identify the presentation features predicting COVID-19 disease diagnoses with high accuracy. Features examined included details of the individuals concerned, e.g., age, gender, observation of symptoms such as fever, cough, difficulty in breathing, etc, history of travel.

We implemented and applied several machine learning algorithms to our collected data and found that the machine learning algorithm performed with the highest accuracy to predict and select features that correctly indicate COVID-19 status for all age groups.

We can use several machine learning algorithm such as logistic regression, neural network, random forest & naive bayes to classify our data. After processing the data we can finally use the ensemble technique to improve the overall performance in our project. Then finally we will get our result in GUI.

1. **INTRODUCTION**

The outbreak of the novel corona virus in early December 2019 in the Hubei province of the People’s Republic of China has spread worldwide. This pandemic continues to challenge medical systems worldwide in many aspects, including sharp increases in demands for hospital beds and critical shortages in medical equipment, while many healthcare workers have themselves been infected.

In this project we will use machine learning approach to identify the symptoms provided by the users. This entire process is done by collecting the data from the user. These data will help to indentify whether any person is suffering from COVID-19 or not, which is based upon some predefined standard symptoms. These symptoms are based on the guidelines given by the World Health Organization (WHO) & the Ministry of Health and Welfare, India.

In this project we will get to know about the dataset contains seven major variables that will bring an impact on whether any person is suffering from corona or not

.

* **Country**: List of the countries a person has visited.
* **Age**: Classification of the age group for each person, based on WHO age standard group.
* **Symptoms**: According to WHO there are five major symptoms such as Fever, Tiredness, Difficulty in Breathing, Dry Cough & Sore throat.
* **Other Symptoms:** Other symptoms include Pain, Nasal Conjection etc.
* **Severity**: The level of severity, Mild, Moderate & Severe.
* **Contact**: Whether person came to contact with a COVID-19 patient.

With all these categorical variables, a combined for each label in the variable will be generated & therefore in total **3,16,800** combinations are created. We need two kinds of data in **csv** (comma separated values) format such as **raw** **data & cleaned data**. In raw data it contains all possible labels of variables, which is used to generate cleaned data. The cleaned data contains all possible from raw data, which can be used for analysis. The cleaned data might contain some dummy variables.

1. **PROBLEM STATEMENT**

Effective screening of SARS-CoV-2 enables quick and efficient diagnosis of COVID-19 and can mitigate the burden on healthcare systems. But due to lack of resources it has become very difficult to identify the symptoms.

With the help of the machine learning approach, the individual person can easily get to know whether they are being infected with corona virus or not. The machine learning model can be used for the analysis & prediction of COVID-19.

1. **AIM**

* The aim is to provide a machine learning model that can easily analysis the symptoms of COVID-19.
* The prediction is performed using the clinical information of the patients.
* The goal is to identify whether a patient can potentially be diagnosed with COVID-19.

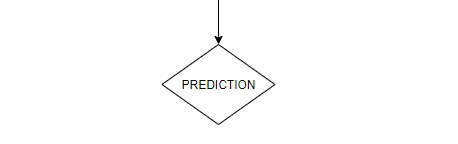
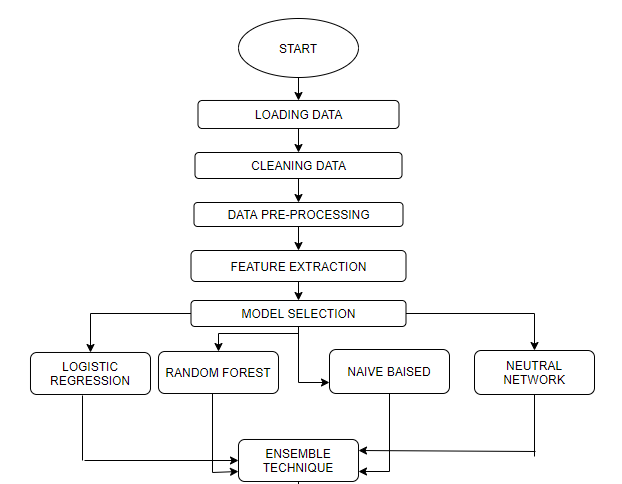
1. **OBJECTIVE**

* The objective of this project is to easily identify the COVID-19 disease by its symptoms.
* It will help to reduce the risk of getting affected by corona virus.
* By using the machine leaning model we will indentify the symptoms of this disease
* Identifying the most suitable machine learning technique for prediction, to perform on clinical reports of patients.

1. **SCOPE & FUTURE WORK**

* This research focuses on development of a machine learning model for predicting COVID-19 in patients.
* We also work to identify the features from the clinical information of patients that would influence the predictive result of COVID-19.
* This project can be used to identify the second wave of the corona virus, if we use sensors & CT-Scanning image data to train our model.

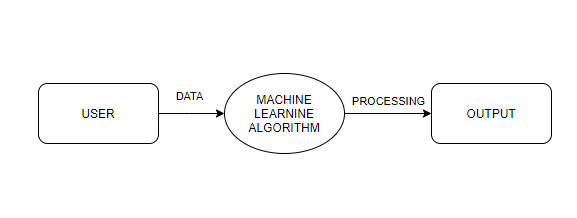
1. **SOFTWARE DESIGN**



NAIVE BAYES

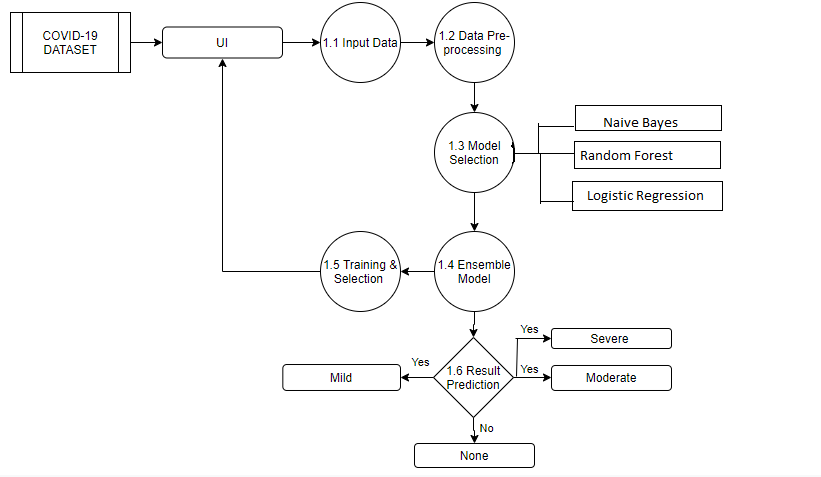
1. **DATA FLOW DIAGRAM**

8 (a) CONTEXT LEVEL DIAGRAM (DFD LEVEL 0)



The context level diagram (Data Flow Diagram Level 0) consist of two external entities the user interface & the output block. The machine learning algorithm represents the use of these algorithms such as logistic regression, random forest, naive bayes & neural network. The output is obtained after processing.

8 (b) CONTEXT LEVEL DIAGRAM (DFD LEVEL 1)



1. **USE CASE DIAGRAM**

**10. CLASS DIAGRAM**

**11. NON FUNCTIONAL REQUIREMENTS**

As the name suggest these are requirement that are not directly interacted with the specific function in this project.

* Performance: The symptoms are taken as input from the users and they are treated as the important feature for analysis.
* Availability: The results are only accurate if the features of the input provided by the users are true and correct.
* Flexibility: It provides the users very comfortable way to get symptoms from user interface and analysis them.
* Learn ability: The software is easy to use & reduced the learning work.

**11. LIMITATIONS**

* This entire project of COVID-19 is based upon the first wave of corona virus i.e. which can be determined by their symptoms.
* This project can only determine the result when the input values are correct.
* The machine learning model can be used to analysis the symptoms of this disease.
* To predict the actual result we have to use the clinical dataset which contains the patient medical reports such as CT scanning & RT-PCR test data.

**12. CONCLUSION**

* In this research, a systematic literature review has been conducted to identify the suitable algorithm for prediction of COVID-19 in patients.
* The selected algorithms were trained with the patient clinical information about the basic symptoms that indicated the infection in a person.

.

* A prediction system that could find the possibility of outbreak of novel diseases that could harm mankind through socio-economic and cultural factor consideration can be developed.
* It is recommended to work on calibrated and ensemble methods that could resolve problems faster with better outcomes than the existing algorithms