



UROP Report

Covid-19 Symptoms Analysis

- Done by

P.Manohar - AP19110010057

V.Akhil - AP19110010118

K.Sravan - AP19110010030

N.Sameer - AP19110010139

Abstract:

The COVID-19 pandemic is one of the disastrous worldwide health crises that occurred in the last century. Its economic, social, and health consequences are only expected to exacerbate, making it one of the biggest global calamities since the 1918 influenza epidemic and the World Wars that happened. Here we will be looking at some of the most essential symptoms for predicting covid-19 as it is not over yet. In the news we can observe that Shanghai, one of the cities in China, is under lockdown due to the corona 4th wave, so we must take the necessary precautions to protect ourselves from the 4th wave. In recent years, the pandemic has created an unprecedented global health crisis. Its economic, social, and health effects are increasing day by day, and it is going to become one of the biggest global disasters ever seen. We are using machine learning algorithms like KNN and Decision trees to perform our analysis.

Keywords: Machine learning, KNN, Decision tree, Prediction, Analysis.

Introduction:

Since the 1918 influenza pandemic, the COVID-19 pandemic has been the most dangerous one that has occurred. It has affected 52+cr people and in which around 62.5 lakh people lost their lives and due to this there are many questions arising such as,

- How did it really started?
- How we can take preventive measures regarding the covid-19
- How to balance both the economy part and public's health
- Reasons behind some countries having low positive cases to death ratio such as NewZealand.

To answer all these questions raised we had tried to develop a model by considering the covid-19 symptoms. The primary objective of Covid-19 symptoms analysis is to analyze due to which symptoms we are getting a covid positive report and the status of the severity of the positive report and this can help people to take preventive measures. The greatest feature of the proposed model is it aims to predict the severity of status of a person when he is tested positive and it also gives based on which symptoms he had been tested positive so that it helps other people such as if they too having the same symptoms then they can go take a covid test which is a good sign.

Here we are training a model with two different algorithms which are K-Nearest Neighbors(KNN) and Decision Trees. Where we will first train the model with these two algorithms and based on the accuracy results obtained we will deploy the model using the best accurate predicted model because we know this covid-19 is one of the scariest health problems we are facing so we should not take any chance of error in this case.

Related Work:

<https://www.nature.com/articles/s41746-020-00372-6#Sec5>

This report used the data the Israeli Ministry of Health publicly released of individuals. The dataset contains initial records, on a daily basis, of all the residents who were tested for COVID-19 nationwide, along with the age, sex and other aspects. They developed a model that predicts COVID-19 test results using eight binary features: sex, age 60 years or above, known contact with an infected individual, and five initial clinical symptoms. The training-validation set consisted of records from 51,831 tested individuals.

This report predicted using a gradient-boosting machine model built with decision-tree base-learners. Missing values were inherently handled by the gradient-boosting predictor. The report estimated differences between models with subsets of the feature space. In addition, they performed plots of the PPV against the sensitivity (precision–recall curve) across different thresholds. Confidence intervals (CI) for the various performance measures were derived through resampling.

Model development:

1. Dataset Description:

Covid is a transmissible disease and the purpose of taking this dataset is to provide symptoms as input and it should be able to predict if covid is possible or not and then it should give us the severity status if the covid is possible. The dataset contains 12 different symptoms that are observed during the covid times and the resulting outcome of that.

	Breathing Problem	Fever	Dry Cough	Sore throat	Running Nose	Asthma	Chronic Lung Disease		Headache	Heart Disease	Diabetes	...	Fatigue	Gastrointestinal
0	Yes	Yes	Yes	Yes	Yes	No	No		No	No	Yes	...	Yes	Yes
1	Yes	Yes	Yes	Yes	No	Yes	Yes		Yes	No	No	...	Yes	No
2	Yes	Yes	Yes	Yes	Yes	Yes	Yes		Yes	No	Yes	...	Yes	Yes
3	Yes	Yes	Yes	No	No	Yes	No		No	Yes	Yes	...	No	No
4	Yes	Yes	Yes	Yes	Yes	No	Yes		Yes	Yes	Yes	...	No	Yes
...

The dataset also consists of some scenarios where we are having a chance of getting a positive result which is covid.

Abroad travel	Contact with COVID Patient	Attended Large Gathering	Visited Public Exposed Places	Family working in Public Exposed Places	Wearing Masks	Sanitization from Market	COVID-19
No	Yes	No	Yes	Yes	No	No	Yes
No	No	Yes	Yes	No	No	No	Yes
Yes	No	No	No	No	No	No	Yes
Yes	No	Yes	Yes	No	No	No	Yes
No	Yes	No	Yes	No	No	No	Yes
...

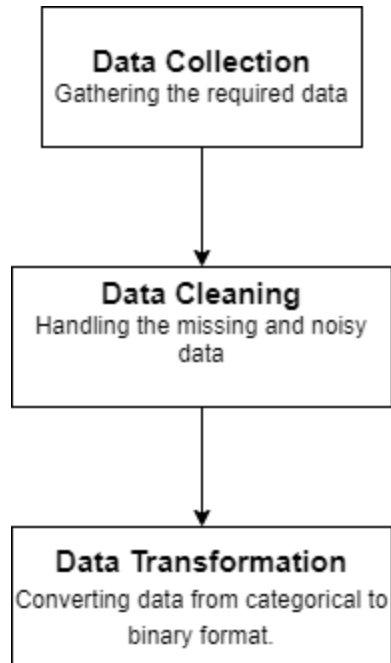
2. Data Preprocessing:

From the available dataset we need to get the best results, so for that we need data preprocessing and as everyone knows the pre-processed data gives a better result than the raw data. Here we performed various preprocessing techniques such as data cleaning and data transformation to get a highly accurate model. This is one of the core steps in our project.

Measures taken while data preprocessing:

- Converting the categorical data into binary form so that the system can read the data.
- Prevention of loss of any data is also taken into consideration.

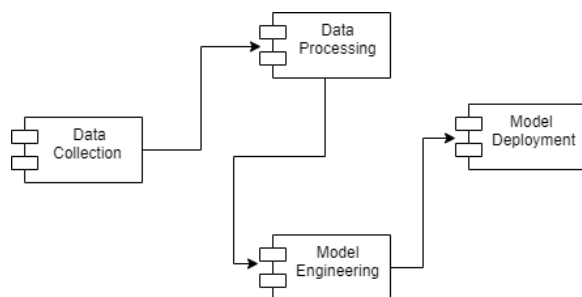
Pre-Processing flow chart:



3. Model Architecture:

The model architecture defines us the overview of the proposed model and it is used to abstract the overall model.

Architecture diagram:



4. Proposed Methodology:

We had designed an automated approach to predict the chances of a person getting covid-19 by applying different algorithm concepts which are K-Nearest Neighbors and Decision trees. So these will consider different parameters from the dataset and then predict the severity status of a covid positive person.

K-Nearest Neighbors:

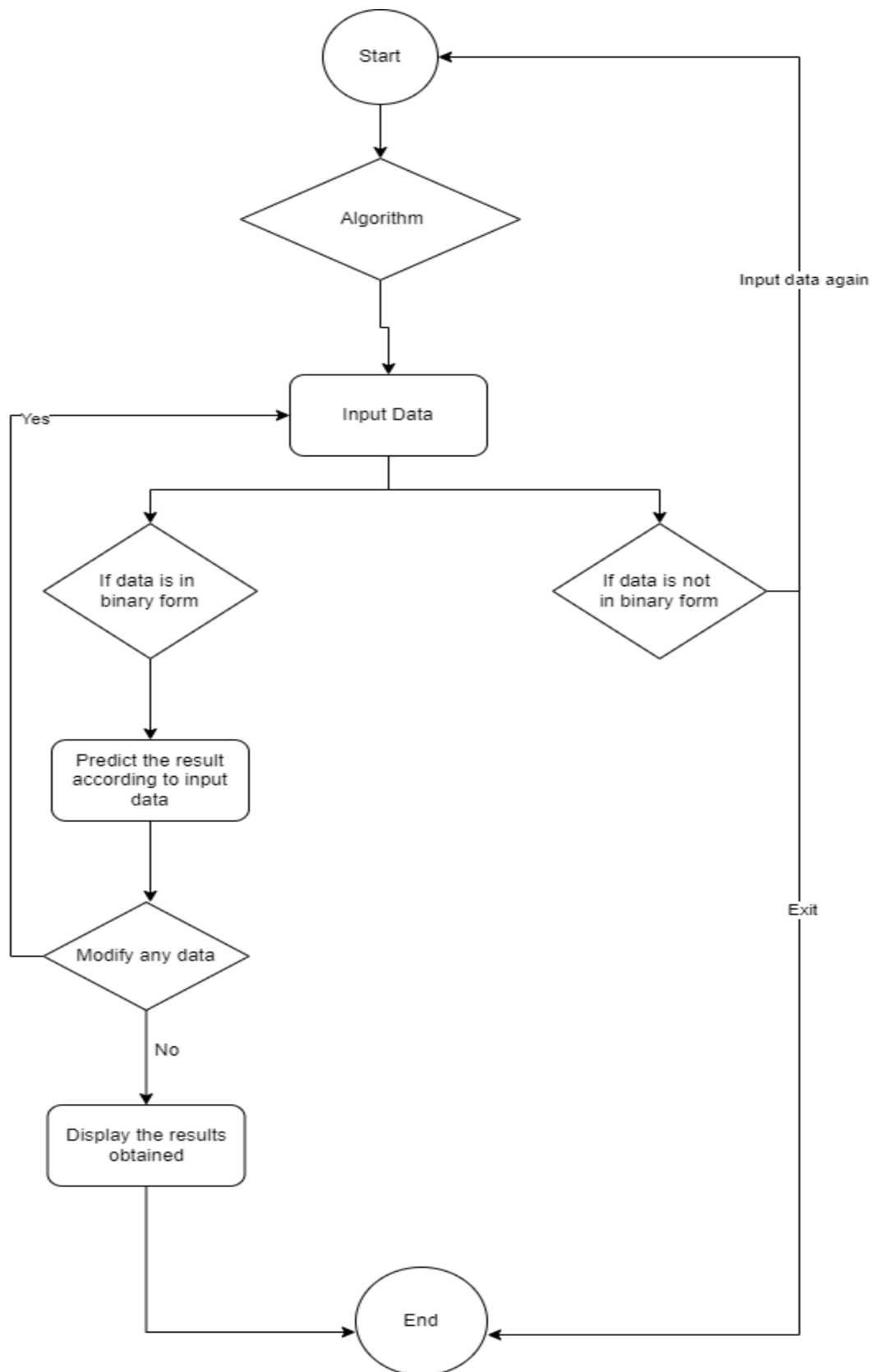
KNN is a supervised machine learning algorithm where we can predict the output, based on the input given and K in KNN means the number of nearest neighbors and we

have used this algorithm because we are having the historic data, so based on that we can predict the severity status of covid positive person by the data collected.

Decision Trees:

Decision tree is a tree-like structure which delivers the consequences based on the input decisions and here we thought this would be one of the best algorithms to implement as severity status depends upon the various input factors and then it gives us the decision.

Flow chart:



Result:

To determine whether or not the person has covid, we had used the Decision tree and Knn algorithm. In comparison to knn, decision tree provided higher accuracy.

Decision Tree report

```
In [98]: from sklearn.metrics import classification_report
print(classification_report(y_test,y_pre))
```

	precision	recall	f1-score	support
0	0.91	0.98	0.95	316
1	1.00	0.98	0.99	1315
accuracy			0.98	1631
macro avg	0.95	0.98	0.97	1631
weighted avg	0.98	0.98	0.98	1631

Knn Report

```
In [99]: print(classification_report(y_test,y_pred))
```

	precision	recall	f1-score	support
0	0.90	0.89	0.89	316
1	0.97	0.98	0.97	1315
accuracy			0.96	1631
macro avg	0.94	0.93	0.93	1631
weighted avg	0.96	0.96	0.96	1631

Conclusion:

The project discusses a few important algorithms used to develop a model to analyze the severity status of the covid of a person based on the symptoms he had. From the analysis results we had obtained, the Decision tree's predictions are more accurate compared to the predictions made from KNN algorithm's model as the accuracy is slightly higher for the decision tree model. Finally, we should not consider this is the end for the one of the scariest health crisis that occurred as we can observe positive cases are globally increasing slowly and we shouldn't take it for granted as each and every individual is not ready for the 4th wave if it comes as everyone are

economically down due to previous pandemic waves. We should take the required precautions and if we have any doubt after observing some symptoms then prefer a covid test.

References:

- <https://www.kaggle.com/>
- <https://link.springer.com/article/10.1007/s41745-020-00200-6>
- <https://www.nature.com/articles/s41746-020-00372-6#Sec5>
- <https://apm.amegroups.com/article/view/45369/html>