Prediction of Covid-19 according to different symptoms

Sarder Iftekhar Ahmed
2018-1-60-181
East West University
email:2018-1-60-181@std.ewubd.edu

Md. Nadim
2018-1-60-161
East West University
email:2018-1-60-161@std.ewubd.edu

Md Mizanur Rahman
2018-1-60-249
East West University
email:2018-1-60-249@std.ewubd.edu

Abstract—**Covid-19 has been a global concern for more than a year and a half. Covid-19 is an infectious disease that is very important to diagnose, but the detection kit is expensive. And because it is time consuming, people are avoiding testing, which is contributing to the spread of infection .Technology advancements have a rapid effect on every field of life; it can be medical field or any other field. Artificial intelligence has shown us better development in health sector by its own decision making and analyzing the data. At this time, our world are going through global pandemic situations. Corona virus has affected most of the country in the world. People have nothing to do except suffering and going through this matter. So, Developing a control system that will detect corona virus will be lifesaving at this moment. we have developed such a tool that will predict covid-19. For making this happen, we used naïve Bayes and a data sheet that we find on internet.**

Index Terms—**Corona Virus, Naïve Bayes, Machine learning.**

I. Introduction

The novel coronavirus emerged in Wuhan city of China in December 2019 and was informed to the World Health Organization (W.H.O) on 31st December 2019. The virus formed a global threat and was named as COVID-19 by W.H.O on 11th February 2020. The COVID-19 comes from the family of viruses including SARS, ARDS. W.H.O declared this outbureak as a public health emergency and stated the following; the virus is being transmitted via the respiratory tract when a healthy person comes in contact with the infected person. The virus may spread between persons through other roots which are currently uncertain. The infected person shows symptoms within 2–14 days, depending on the incubation period of the middle east respiratory syndrome (MERS), and the severe acute respiratory syndrome (SARS). According to WHO the signs and indications of mild to moderate cases are dry cough, fatigue and fever while as in severe cases dyspnea (shortness of breath), Fever and tiredness may transpire. The persons having other diseases like asthma, diabetes, and heart disease are more exposed to the virus and maybecome severely ill. The person is diagnoses based on symptoms and his travel history. Vital signs are being experimental keenly of the client having symptoms. By washing hands regularly with soap for at least 20 seconds and avoiding close body contact with others by keeping the distance of about 3 meter may reduce the chances of getting affected by this virus. While sneezing, covering the mouth and nose with the help of throwaway tissue

and avoiding the contact with the nose, ear and mouth can help in its prevention. Till 23th of May 2021, almost 167 million confirmed cases of coronavirus are spotted around the globe. Almost 3.4 million persons have died and 148 million persons have recovered from this deadly virus. Figure 1 displays the worldwide data regarding coronavirus. Less time and man- power in testing have also given rise to this disease as we lack the medical resources due to pandemic. Since thousands of people are being tested positive every day around the globe, it is not possible to test all these people who show symptoms Apart from clinical procedures, machine learning provides a lot of support in detecting the disease with the help of image and textual data. Machine learning can be used for the identification of novel coronavirus. It can also predict the nature of the virus across the globe. However, machine learning requires a large amount of data for classifying or foreseeing diseases. Supervised machine learning algorithms need marked data for classifying the text or image into different categories. From the past decade, a large amount of growth is being made in this area for resolving some critical projects. Recent pandemic has attracted many researchers around the globe to resolve this problem. So, this is no surprise that covid-19 is a big concern for us , people are implementing many approaches and doing many research to prevent and detecting corona virus. In this project, we will discuss about the detection of corona virus based on different symptoms using Naïve Bayes classifier.

II. Related Work

Machine learning and natural language processing uses big data-based replicas for pattern recognition, explanation, and prediction. NLP has expanded much interest in recent years, mostly in the field of text analytics, Classification is one of the foremost tasks in text mining and can be accomplished using different algorithms. Kumar et al performed a SWOT analysis of various supervised and unsupervised text classification algorithms for mining the unstructured data [1]. The various applications of text classification are sentiment analysis, fraud detection, and spam detection etc. Opinion mining is majorly being cast-off for elections, advertisement, business etc. Verma et al analyzed Sentiments of Indian government projects with the help of the lexicon-based dictionary [2]. The machine learning has changed the viewpoint of diagnosis by giving

great results to diseases like diabetes and epilepsy. These purposes can be helpful to diagnose and predict COVID-19. Firm and exact diagnosis of COVID-19 can save millions of lives and can harvest a massive amount of data on which a machine learning (ML) models can be trained. ML may provide useful input in this favor, in particular in making diagnoses based on clinical text, radiography Images etc. According to Bullock et al, Machine learning and deep learning can substitute humans by giving an accurate diagnosis. The perfect diagnosis can save radiologists' time and can be cost-effective than normal tests for COVID- 19 [3]. X-rays and computed tomography (CT) scans can be used for training the machine learning model. Several initiatives are ongoing in this regard. Wang and Wong developed COVID-Net, which is a deep convolutional neural network, which can diagnose COVID-19 from chest radiography images [4]. Once the COVID-19 is detected in a person, the query is whether and how intensively that person will be affected. Not all COVID-19 positive patients will need rigorous consideration. Being able to prognosis who will be affected more severely can help in directing assistance and planning medical resource allocation and application. Jiang et al proposed a machine learning model that can foresee a person affected with COVID-19 and has the possibility to develop acute respiratory distress syndrome

(ARDS) [5]. The projected model resulted in 80% of accuracy. The samples of 53 patients were used for training their model and are limited to two Chinese hospitals. Since less work is being done on detecting corona virus, we used machine learning Naïve Bayes algorithm to detect corona virus based on different symptoms.

III. METHODOLOGY

Our proposed project methodology consists of two different steps one is data collection and another one is machine-learning classification. In the given diagram below, we try to show our entire methodology. At machine learning classification, we will discuss about which machine-learning algorithm we are going to use for this proposal project.
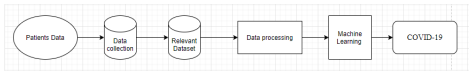

Fig. 1. Control Flow Diagram

A. Data Collection

As the World Health Organization declared health emergency because of Corona virus pandemic

around the globe. All the researcher and hospitals around the world has given open access to the data about this digester. As a result, we managed to collect our data from open-source data repository on the internet. In which, 1001 rows of data is sorted which have shown symptoms and results of corona virus affected peoples. Data consists of about different attributes such as serial number, Gender, Temperature, cough, Tiredness difficulty, Loss of taste, Sore throat and result of covid-19.

B. Machine-Learning Classsification

In this project, we used Naïve Bayes classifier for corona virus results prediction. A Naïve Thomas Bayes classifier could be a supervised machine- learning formula that utilize Bayes Theorem. In which, we predict that feature are statically non- depended. It actually depends on the Naïve prediction that input variable are freelance of every alternative. There is no thanks to acknowledge something concerning alternative variable once given a further variable. In any case of this prediction, it has justified itself to be a classifier with good results. Naïve Bayes classifier depends on the Bayes Theorem, which is based on conditional probability, the likelihood that an event will occur given that another event has already took place. Basically, this Theorem help a hypothesis to be modified all the time as new evidence is introduced, the equation give below expresses Bayes theorem within the language of probability.

Bayes theorem gives a way of calculating posterior probability $P(A|B)$ from P(A), P(B) and $P(B|A)$. Look at the equation 2

Fig. 2.

Bayes theorem above,

$P(A|B)$ is the posterior probability of class (A, target) given predictor (B, attributes). P(A) is the prior probability of class.

$P(B|A)$ is the likelihood, which is the probability of predictor given class. P(B) is the prior probability of predictor.

Let us understand it using our example; below we have a data of covid-19 results based on different symptoms. Now, we need to classify whether people are covid-19 affected or not based on the data of the table.

Step-1: convert the data into a frequency table.

Step-2: create likelihood data table by finding the probabilities like overcast probability. Step-3: At last, we use Naïve Bayes equation to calculate the posterior probability for each class. The class with the most superior posterior probability is the outcome of the prediction.

IV. RESULTS DISCUSSION

we used a windows system with 12 GB Ram and 3.4 GHz processor for performing this work. Jupiter notebook tools

1	Serial No	Gender	Temperat	Cough	Tiredness	Difficulty_	Loss_of_ta	Sore_thro	Covid-19
2	10001	female	104	dry	severe	yes	yes	yes	positive
3	10002	female	100	dry	severe	no	no	no	negative
4	10003	female	98	wet	severe	no	no	no	negative
5	10004	female	105	wet	slight	yes	yes	no	negative
6	10005	female	101	dry	slight	no	yes	yes	positive
7	10006	male	103	dry	severe	yes	yes	yes	positive
8	10007	male	96	no	slight	no	no	no	negative
9	10008	male	99	wet	slight	no	no	yes	negative
10	10009	male	98.5	wet	no	no	no	no	negative
11	10010	female	102.5	no	slight	yes	yes	no	positive
12	10011	male	102	no	severe	yes	yes	yes	positive
13	10012	male	99	no	no	no	no	yes	negative
14	10013	male	101	no	slight	no	yes	yes	positive
15	10014	female	101.5	no	severe	yes	yes	no	positive
16	10015	male	102.5	wet	slight	no	yes	yes	negative

Fig. 3. data Set

used for running our machine learning algorithm with the help of different libraries for improving our results.

A. Preprocessing

The dataset had to be processed in several ways. First, we have to find the unique attribute and total value of deciding factor to find the number of positive and negative people. Then, we find the probabilities of both positive and negative people affected people from data sheet.

B. Testing And Evaluation

As we said before, we have used Naïve Bayes classifier in this project and our results pretty much as our expectations. We have shown a sample of our results below this.


```
Temperature: 100
Cought status('dry','wet','no'): dry
Tiredness status('severe','slight','no'): slight
Difficulty Breathing status('yes','no'): no
Loss of taste or smell status('yes','no'): yes
Score throat status('yes','no'): no
no p 0.6264501160092807 n 0.6274165202108963
yes p 0.6102088167053364 n 0.5817223198594025
no p 0.6125290023201856 n 0.6537785588752196
slight p 0.419953596287703 n 0.4446397188049209
dry p 0.43387470997679817 n 0.37434094903339193
100.0 p 0.4779582366589327 n 0.4780316344463972
Actual Probability of Postive: 0.4485915343557016 Actual Probability of Negative: 0.5514084656442982
Covid-19: Negative
```

Fig. 4. result

V. CONCLUSION

We used 1001 rows data reports, which are characterized in different classes. The machine learning algorithms are used for finding the probabilities of covid-19. After performing classification, it was discovered that Naïve Bayesian classifier gives excellent results. Various other machine learning algorithms that displayed better results were random forest, stochastic gradient boosting, decision trees and boosting. The efficiency of models can be enhanced by increasing the amount of data. Also, the disease can be classified on the gender-based such

that we can get info about whether the males are affected more or females. More featured engineering is needed for better results and deep learning approach can be used in future.

REFERENCES

[1] A. Kumar, V. Dabas, and P. Hooda, "Text classification algorithms for mining unstructured data: a swot analysis," *International Journal of Information Technology*, vol. 12, no. 4, pp. 1159–1169, 2020.
[2] P. Verma, A. Khanday, S. T. Rabani, M. H. Mir, and S. Jamwal, "Twitter sentiment analysis on indian government project using r," *Int J Recent Technol Eng*, vol. 8, no. 3, pp. 8338–41, 2019.
[3] J. Bullock, A. Luccioni, K. H. Pham, C. S. N. Lam, and M. Luengo-Oroz, "Mapping the landscape of artificial intelligence applications against covid-19," *Journal of Artificial Intelligence Research*, vol. 69, pp. 807–845, 2020.
[4] L. Wang, Z. Q. Lin, and A. Wong, "Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images," *Scientific Reports*, vol. 10, no. 1, pp. 1–12, 2020.
[5] W. Xu, N.-N. Sun, H.-N. Gao, Z.-Y. Chen, Y. Yang, B. Ju, and L.-L. Tang, "Risk factors analysis of covid-19 patients with ards and prediction based on machine learning," *Scientific reports*, vol. 11, no. 1, pp. 1–12, 2021.