



# IJRASET

International Journal For Research in  
Applied Science and Engineering Technology



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 10    **Issue:** XII    **Month of publication:** December 2022

**DOI:** <https://doi.org/10.22214/ijraset.2022.48002>

[www.ijraset.com](http://www.ijraset.com)

Call:  08813907089

E-mail ID: [ijraset@gmail.com](mailto:ijraset@gmail.com)

# Disease Prediction System using naïve bayes

Sameer Meshram<sup>1</sup>, Shital Dongre<sup>2</sup>, Triveni Fole<sup>3</sup>

<sup>1,2,3</sup>Department of Information Technology Vishwakarma Institute of Technology

**Abstract:** *Accurate and on-time analysis of any health-related problem is vital for the prevention and treatment of the illness. The standard way of diagnosis might not be sufficient. Developing a diagnosis system with machine learning (ML) algorithms for prediction of any disease can help in a very more accurate diagnosis than the traditional method. The proposed model is an Disease Prediction System with the help of machine learning algorithm Naïve Bayes which takes the symptoms as the input and it gives the output as predicted disease. It results in saving time and also makes it easy to induce a warning about your health before it's too late. By using this model anyone can get the result as predicted disease by simply given the symptoms as input. The accuracy of this model is more than existing models.*

**Keywords:** *Disease prediction, more accuracy, symptoms, naïve bayes.*

## I. INTRODUCTION

It is observed that near about 67-70 % of the population of India were affected by some common diseases like flu, cold cough, diarrhoea etc in each couple of months. Number of peoples even don't realize that some common symptoms can give rise to major diseases. It is estimated that 20-25 % of population results to very bad situation and deaths just because of avoiding some common symptoms. Due to such concern catching or predicting such small diseases at very early stage is able to resist unwanted casualties. The existing system has the capability to just predict any specific disease which will bound its limits.

The main objective of the system is to predict such diseases that if we avoid or when unchecked can result into fatal situations. By embedding machine learning algorithms like naïve Bayes part of supervised learning algorithm. This model can generate a prediction of the most possible disease matching with symptoms which is trained using machine learning mechanism. The proposed model provides quick medical diagnosis to users. It is very easy to an user to get to know the predicted chronic disease by simply giving the symptoms as input.

## II. LITERATURE SURVEY

In the proposed model "Disease Prediction Using Machine Learning" the author has discussed that the Disease prediction based on history of patient treatment and health data by applying machine learning and data processing techniques with the ongoing struggle for the past decades. Also in the number of cases they are applying the data processing onto the medical profiles for prediction of specific diseases. These approaches tried to predict the occurrence of disease. [1]

In the proposed model "A Smart Health Prediction Using Data Mining" the information about mining techniques like association rule mining and classification, clustering and also proposed a model which makes use of various techniques like machine learning, artificial Intelligence, management techniques for the extraction of new patterns from large datasets and therefore the knowledge related to these patterns. By using this technique data can be extracted automatically or semi automatically. The various parameters included in data processing are clustering, forecasting, path analysis and predictive analysis. [2]

In the paper "General Disease Prediction System" the author has discussed about data processing techniques and ID3 decision tree algorithms. Within the project, the disease prediction system will perform data processing in its preliminary stages, also the system are trained using machine learning and data processing to form the prediction for the final and more commonly occurring disease that when unchecked can transform fatal disease. [3]

In the proposed mode "Smart E-Health Prediction System Using Data Mining" it discusses the Bayesian statistics and posterior distribution the puzzles that are solved highly and also told us about data processing techniques like association rule mining, classification, clustering. Also Bayesian statistics successfully can be applied on different fields like science, economic. In the fields like medical fields, the foreign students have solved some medical issues that are laborious to be settled in classic statistics by classification of Bayesian. [4]

In the paper "Medi-Insight: A Smart Health Prediction System" the author proposes an answer for identifying diseases supported symptoms Information technologies which are being increasingly implemented in healthcare organizations.

Their system can able to predict a person’s lifetime diseases are going to be able to warn the person to possess early medical services or to manage and stop exceptional situations in life with the assistance of Multinomial Naïve Bayes Algorithm. The Author used Multinomial Naive Bayes which may be a variant of Naive Bayes algorithm accustomed implement our disease prediction model. Multinomial naive Bayes assumes to own feature vector where each element represents the amount of times it appears (or, very often, its frequency). this method is additionally very efficient in linguistic communication processing or whenever the samples are composed ranging from a standard dictionary.[5]

In the proposed model, K means clustering for clustering algorithm and later cluster the data an approach of KNN is used here the dataset of diabetes and liver disorder are used for performing test operation onto the defined methods. the concept of fuzzy logic is used for improving the accuracy of classification .[6]

In the proposed model , combination Bayesian and KNN(K- Nearest Neighbour) algorithm are used .here a dataset of diabetes patient used that gives an analysis of diabetes malady[7]

In the proposed model, diagnosis of heart disease based on the previous information and related data. Here the concerned risk factor of heart disease predicted using the naïve bayes. The parameters considered are sex, gender, cholesterol, bp, blood sugar etc. They are act as input for the prediction.[8]

In the proposed model, machine learning algorithms for the prediction of chronic diseases are used. Here the model is tried as experiment over the data which is collected from the real world hospital in 2013-2015. they used here the convolutional neural network approach for multiple disease predictions reached the prediction accuracy of 90%. [9]

In the proposed model, for the prediction of multiple diseases like diabetes analysis, heart disease, diabetes etc an API named as Flask is used. The parameters here taken into consideration are bmi, sex, glucose, diabetes pedigree function, heart rate/pulse rate, serum creatinine, potassium Glasgow Coma scale .[10]

### III. PROPOSED MODEL

As we were thinking on how and what we can do in the disease prediction field, So we Developed Web-app which predicts disease on basis of user’s symptoms. So in this project we are predicting a disease which an individual is affected by depending upon the symptoms patient is suffering .In the proposed model , system is asking for symptoms and evaluating those symptoms by using the Naïve Bayes algorithm which works robust and faster with datasets having small in size also it gives the nice accuracy. Naive Bayes algorithm which learns the probability of an object with certain features belonging to a specific group/class. for example, if we are trying to identify a fruit based on its taste ,shape and color then an orange colored, tangy fruit and spherical would presumably be an orange. Such properties can be used individually contribute to the probability that this fruit is an orange. Due to these reason it is referred to as “Na-ive”. As for the “Bayes” part , it refers to statistician and philosopher Thomas Bayes and also the theorem named after him , Bayes’ theorem , which is that the base for Naïve Bayes.

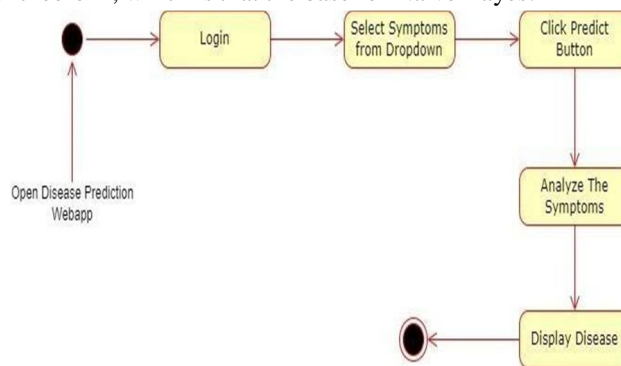


Fig: State Diagram

### IV. METHODOLOGY

#### A. Algorithm

Step 1: Let Say D can be represented as training dataset every record can be denoted as n-dimensional attribute vector, thus we can write this as  $X=(x_1,x_2,x_3,x_4,\dots,X_n)$  this is responsible to predict n attributes of n measurements ( $B_1$  to  $B_n$ ).

Step 2: Lets take m number of class for operation of prediction it can written as ( $C_1,C_2,C_3,\dots,C_m$ )

By applying the Naive Baye's theorem:

$$P(C_i | X) = \frac{P(X | C_i) * P(C_i)}{P(X)}$$

$$P(X)$$

Step 3: Here for every class P(X) is constant, therefore P(X|C)\*P(Ci) compulsorily be maximized.

Step 4: After,  $P(X | C_i) = P(x_1 | C_i) * P(x_2 | C_i) * \dots * P(x_m | C_i)$

Step 5: for X class prediction, P(X|Ci)P(Ci) is need to find outfor each Ci class

Here classifier i.e naïve bayes is able to predict the class labelof X=Ci class

If,

$$P(X|C_i) P(C_j) > P(X|C_j)P(C_j)$$

### B. Steps Of Model Building

- 1) *Collecting Data:* Data from excel, access, text files etc., this step (gathering past data) can be act as a inspiration of the longer term learning. The higher the variability, volume and density of relevant data, better the aspects of learning becomes.
- 2) *Pre-processing the Data:* This steps used to re- move outliers and missing data to form the clean data. Exploratory analysis is maybe one method to review the nuances of the information in details thereby burgeoning the nutritional content.
- 3) *Training a Model:* This step involves choosing the acceptable algorithm and representation of knowledge within the type of the model. The data that we get after removing outliers and cleaned data is split into two parts – train and test (proportion reckoning on the prerequisites); the primary part (training data) is employed for developing the model. Another step(test data) is employed.

❖ Following are the Parameters that taken into consideration for training purpose are shown below:

Itching, skin\_rash, nodal\_skin\_eruptions, continuous\_sneezing, shivering, chills, joint\_pain, stomach\_pain, acidity, ulcers\_on\_tongue, muscle\_wasting, vomiting, burning\_micturition, spotting\_urination, fatigue, weight\_gain, anxiety, cold\_hands\_and\_feets, mood\_swings, weight\_loss, restlessness, lethargy, patches\_in\_throat, irregular\_sugar\_level, cough, high\_fever, sunken\_eyes, breathlessness, sweating, dehydration, indigestion, headache, yellowish\_skin, dark\_urine, nausea, loss\_of\_appetite, pain\_behind\_the\_eyes, back\_pain, constipation, abdominal\_pain, diarrhoea, mild\_fever, yellow\_urine, yellowing\_of\_eyes, acute\_liver\_failure, fluid\_overload, welling\_of\_stomach, swelled\_lymph\_nodes, malaise, redness\_in\_eyes, redness\_of\_eyes, sinus\_pressure, runny\_nose, congestion, chest\_pain, weakness\_in\_limbs, fast\_heart\_rate, pain\_during\_bowel\_movements, pain\_in\_anal\_region, bloody\_stool, irritation\_in\_anus, neck\_pain, dizziness, cramps, bruising, obesity, swollen\_legs, swollen\_blood\_vessels, severe\_pain\_in\_chest, puffy\_face\_and\_eyes, enlarged\_thyroid, brittle\_nails, swollen\_extremities, excessive\_hunger, extra\_marital\_contacts, dryness\_and\_tingling\_lips, slurred\_speech, knee\_pain, hip\_joint\_pain, muscle\_weakness, stiff\_neck, swelling\_joints, movement\_stiffness, spinning\_movements, loss\_of\_balance, unsteadiness, weakness\_of\_one\_body\_side, loss\_of\_smell, bladder\_discomfort, foul\_smell\_of\_urine, continuous\_feel\_of\_urine, passage\_of\_gases, internal\_itching, toxic\_look(typhos), depression, irritability, muscle\_pain, altered\_sensorium, red\_spots\_over\_body, belly\_pain, abnormal\_menstruation, dischromic\_patches, watering\_from\_eyes, increased\_appetite, polyuria, family\_history, mucoid\_sputum, rusty\_sputum, lack\_of\_concentration, visual\_disturbances, receiving\_blood\_transfusion, receiving\_unsterile\_injections, coma, stomach\_bleeding, distention\_of\_abdomen, history\_of\_alcohol\_consumption, fluid\_overload, blood\_in\_sputum, prominent\_veins\_on\_calf, palpitations, painful\_walking, pus\_filled\_pimples, blackheads, scurring, skin\_peeling, silver\_like\_dusting, small\_dents\_in\_nails, inflammatory\_nails, blister, red\_sore\_around\_nose, yellow\_low\_crust\_ooze, prognosis.

❖ Following are the disease our model is predicting: Fungal, infection, Allergy, GERD, Chronic, cholestasis, Drug, Reaction, Peptic, ulcer, disease, AIDS, Diabetes, Gastroenteritis, Bronchial, Asthma, Hypertension, Migraine, Cervical, spondylosis, aneurysm (brain hemorrhage), Jaundice, Malaria, Chicken pox, Dengue, Typhoid, hepatitis, A, Hepatitis, B, Hepatitis C, Hepatitis, D, Hepatitis, E, Alcoholic hepatitis, Tuberculosis, Common, Cold, Pneumonia, Dimorphic, hemorrhoids (piles), Heart, attack, Varicose veins, Hypothyroidism, Hyperthyroidism, Hypoglycemia, Osteoarthritis, Arthritis, (vertigo) Parosmia, Positional, Vertigo, Acne, Urinary, tract infection, Psoriasis, Impetigo Fungal infection

- Evaluating the Model:** For determining the performance we have to denote the TP,FP,TN and FN considered as true positive i.e the instances are correctly predicted. False negative i.e instances are incorrectly predicted as not required,false positive i.e instanced predicted are incorrect as required, true positive i.e instances predicted are correct as required.

$$\text{Precision} = \frac{TP + TN}{TP + TN}$$

$$TP + TN + FP + FN$$

$$\text{Accuracy} = \frac{TP}{TP + FP}$$

- Improving the Performance:** In this step it needs to choose different model altogether or increasing the number of variables to improve efficiency. Due to these reason most of time need to spent in the data preparation and data collection.

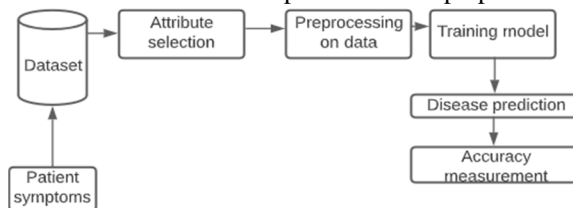


Fig: flow diagram

## V. TOOLS AND TECHNOLOGY USED

### A. Front-end

We have used HTML and CSS for front-end development with django framework which is a Python-based free and open-source web framework that follows the model template views architectural pattern and give us great advantage for integrating with python.

### B. Back -end

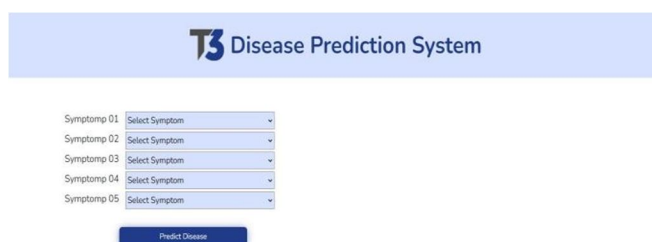
We have used Python in back-end which provided us various libraries such as numpy, pandas and sklearn which helped us to implement this process in an easy manner.

### C. Results

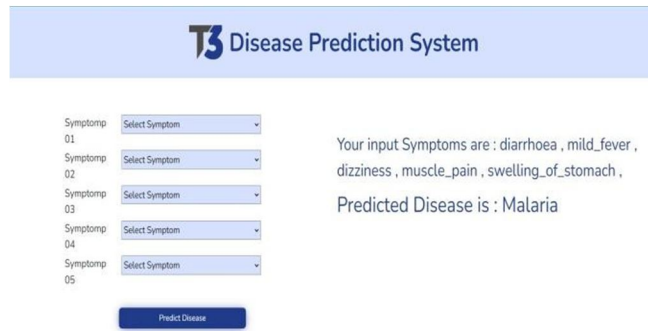
#### 1) Homescreen



#### 2) Dashboard of symptoms



3) Predicted Result



D. Advantages

- High Accuracy compared to existing models.
- Computational Speed is High.
- Multiple disease prediction is possible.

Comparison of existing models with proposed model

Related work	Accuracy	Diseases able to predict
Multi Disease Prediction Model by using Machine Learning and Flask API	91 %	30-40
Diabetes Disease Prediction Using Data Mining	70%	1
Disease Prediction by Machine Learning over Big Data from Healthcare Communities	94.8%	40-50
Disease Prediction System using Naïve Bayes(our proposed model)	98.2%	100-150

VI. CONCLUSION

The main motivation is to facilitate coordinated systems and well-informed health care systems. The proposed model helps any user to provide consultancy of his concerned symptoms. Depending on the symptoms given as input or in case nothing is given the model is capable of predicting the matched disease. The model having the capability of predicting the probability of diseases which can occur results in reduction of mortality ratio.

REFERENCES

- [1] Akash C. Jamgade, Prof. S. D. Zade, International Research Journal of Engineering and Technology, March 2019
- [2] Prof. Krishna Kumar Tripathi, Shubham Jawadwar, Siddhesh Murudkar, Prince Mishra, "A Smart Health Prediction Using Data Mining" International Research Journal of Engineering and Technology (IR- JET), April 2018.
- [3] Shratik J. Mishra 1, Albar M. Vasi 2, Vinay S. Menon3, Prof. K. Jayamalani4, " General disease prediction system" , International Research Journal of Engineering and Technology (IRJET), March 2018
- [4] G.Pooja reddy, M.Trinath basu, K.Vasanthi, K.Bala Sita Ramireddy, Ravi Kumar Tenali, "Smart E-Health Prediction System Using Data Mining", International Journal of Innovative Technology and Exploring Engineering(IJITEE), April 2019.
- [5] Yashaswi G Sagar1, Sahana Gajanana Acharya2, Vishal S Chincholi3, Riyal Vivek A4, Swetha P M5," Medi-Insight: A Smart Health Prediction System", International Research Journal of Engineering and Technology (IRJET), June 2021.
- [6] [4] Naganna Chetty, Kunwar Singh Vaisla, Nagamma Patil," An Improved Method for Disease Prediction using Fuzzy Approach", ICACCE, 2015.
- [7] Deeraj Shetty, Kishor Rit, Sohail Shaikh, Nikita Patil," Diabetes Disease Prediction Using Data Mining", ICIECS, 2017.
- [8] Akkem Yaganteeswarudu," Multi Disease Prediction Model by using Machine Learning and Flask API", IEEE, July 2022.
- [9] Min Chen, Yixue Hao, Kai Hwang, Fellow, Lu Wang, and Lin Wang," Disease Prediction by Machine Learning over Big Data from Healthcare Communities", IEEE, April 2016.
- [10] Anjan Nikhil Repaka, Sai Deepak Ravikanti, Ramya G Franklin," Design And Implementing Heart Disease Prediction Using Naives Bayesian", IEEE, June 2019.



10.22214/IJRASET



45.98



IMPACT FACTOR:  
7.129



IMPACT FACTOR:  
7.429



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24\*7 Support on Whatsapp)