



IJIRCCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 6, June 2021

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 7.542



9940 572 462



6381 907 438



ijircce@gmail.com



www.ijircce.com

Advanced Disease Prediction Using Machine Learning

Aditya Sharma, Anurag Mahajan, Saurabh Chaudhari, Bhagyashree Indalkar, Prof.Vajid.N. Khan

Department of Computer Engineering, Jaywantrao Sawant College of Engineering, Maharashtra, India

Department of Computer Engineering, Jaywantrao Sawant College of Engineering, Maharashtra, India

Department of Computer Engineering, Jaywantrao Sawant College of Engineering, Maharashtra, India

Department of Computer Engineering, Jaywantrao Sawant College of Engineering, Maharashtra, India

Department of Computer Engineering, Jaywantrao Sawant College of Engineering, Maharashtra, India

ABSTRACT: In today's era, each and every human-being on earth depends on medical treatment and medicines. Every day we can hear some new diseases or new symptoms of the existing disease being discovered. But with the growing number of diseases and their symptoms, everyone cannot manage to be updated with it. So to deal with such situations, we are developing an application "Advance Disease Prediction Using Machine Learning" which has a list of large number of diseases, their symptoms, and their treatment required. One major problem in today's world is hike in Doctor's fee. So, the middle class and lower class people are unable to afford for the fee and treatment charges. The application is developed taking this fact in mind. Using this application, one can easily find what disease he/she is infected with by simply inputting the symptoms faced. There are some other features such as inquiring about the diseases etc.

KEYWORDS: Decision tree, Random Forest, Naive Bayes Algorithm, KNN Algorithm, Application, Data Mining, Symptoms.

I. INTRODUCTION

Now-a-days, people face various diseases due to the environmental condition and their living habits. So the prediction of disease using patient treatment history and health data by applying data mining and machine learning techniques is ongoing struggle for the past decades. The recent success of deep learning in disparate areas of machine learning has driven a shift towards machine learning models that can learn rich, hierarchical representations of raw data with little preprocessing and produce more accurate results. The main focus is on to use machine learning in healthcare to supplement patient care for better results. Machine learning has made easier to identify different diseases and diagnosis correctly. Predictive analysis with the help of efficient multiple machine learning algorithms helps to predict the disease more correctly and help treat patients. The healthcare industry produces large amounts of healthcare data daily that can be used to extract information for predicting disease that can happen to a patient in future while using the treatment history and health data. For the disease prediction, we have used K-Nearest Neighbor (KNN), Random Forest Algorithm, Decision tree and Naïve Bayes algorithm machine learning algorithm for accurate prediction of disease. After general disease prediction, this system is able to give the risk associated with general disease which is lower risk of general disease or higher.

II. LITERATURE SURVEY

The paper presented by Dr. Mahboob Khan (PHD) Implements Smart Health Prediction Using Data-Mining. Data Mining is a technology which uses already existing data in the database to manipulate results. It also uses data mining and Database management system to extract knowledge from the large set of data sets. The database is fed with the list of various diseases, information about those diseases, their symptoms, and medicines. User is expected to input the symptoms he/she deals with. The system processes all the symptoms to search for different diseases associated with it and output the diseases which are most probable.

A Survey of Health Care Prediction Using Data Mining by Sujatha, Sumathy, AnithaNithya suggests Data Mining as one of the most motivating area that is becoming popular in health organization .The actual task of data mining is to extract data by automatic or semi-automatic means. Different areas of mining include clustering, forecasting,path analysis.

A paper by Daniel Lowd and Pedro Domingos on Naïve Bayes model for probability estimation, aimed to show that for a wide range of datasets, Naïve Bayes models have accuracy and less learning time compared to other Bayesian networks. The magnitude of order of Naïve Bayes inference is faster than Bayesian network inference.

III.SYSTEMARCHITECHTURE

Disease prediction using machine learning predicts the presence of the disease for the user based on various symptoms and the information the user gives such as sugar level, haemoglobin level and many more such general information through the symptoms. The architecture of the system disease prediction using machine learning consist of various datasets through which we will compare the symptoms of the user and predicts it, then the datasets are transformed into the smaller sets and from there it gets classified based on the classification algorithms later on the classified data is then processed into the machine learning technologies through which the data gets processed and goes in to the disease prediction model using all the inputs from the user that is mentioned above. Then after user entering the above information and overall processed data combines and compares in the prediction model of the system and finally predicts the disease. An architecture diagram is a graphical representation of a set of concepts, that are part of an architecture, including their principles, elements and components. The diagram explains about the system software in perception of overview of the system.

IV. METHODOLOGY

Advance disease prediction using Machine learning is an application which aims to predict the disease on the basis of the symptoms. Patient will specify the symptoms caused due to his illness. System will ask some questions regarding his illness and it will predict the disease from which he may be suffering using an efficient data mining algorithm. Patient can also search about the disease from which he may be suffering.

4.1 Database Collection-

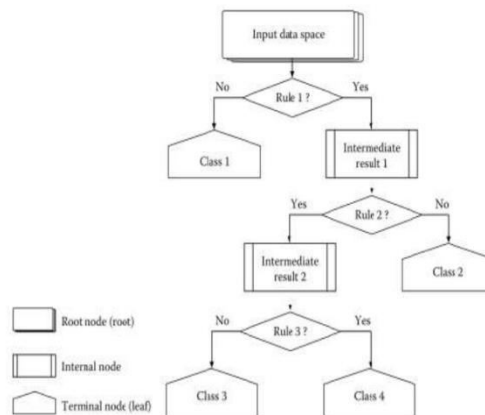
Data collection has been done from the internet to identify the disease here the real symptoms of the disease are collected. Dataset for this project was collected from a study of university of Columbia performed at New York Presbyterian Hospital during 2004. Link of dataset is given below.
<http://people.dbmi.columbia.edu/~friedma/Projects/DiseaseSymptomKB/index.htm>

4.2 Models-

There are four different kind of models present in our project to predict the disease these are

- Decision tree
- Random forest tree
- Gaussian Naïve Bayes
- KNN

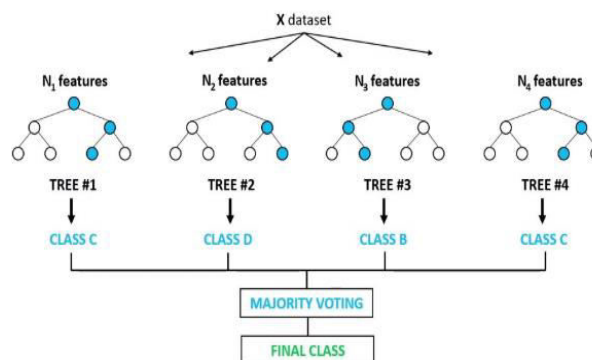
Decision tree- It is classified as a very effective and versatile classification technique. It is used in pattern recognition and classification for image. It is used for classification in very complex problems due to its high adaptability. It is also capable of engaging problems of higher dimensionality. It mainly consists of three parts root, nodes and leaf. Roots consists of attribute which has most effect on the outcome, leaf tests for value of certain attribute and leaf gives out the output of tree. Decision tree is the first prediction method we have used in our project. It gives an accuracy of ~95%.



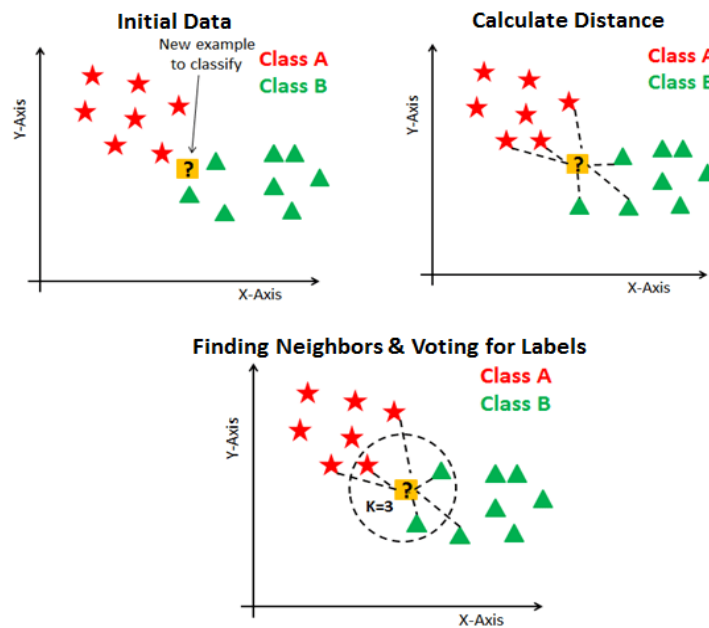
Random Forest Algorithm- It is a supervised learning algorithm used for both classification and regression. This algorithm works on 4 basic steps –

1. It chooses random data samples from dataset.
2. It constructs decision trees for every sample dataset chosen.
3. At this step every predicted result will be compiled and voted on.
4. At last most voted prediction will be selected and be presented as result of classification. In this project we have used random forest classifier with 100 random samples. The result given is ~95% accuracy.

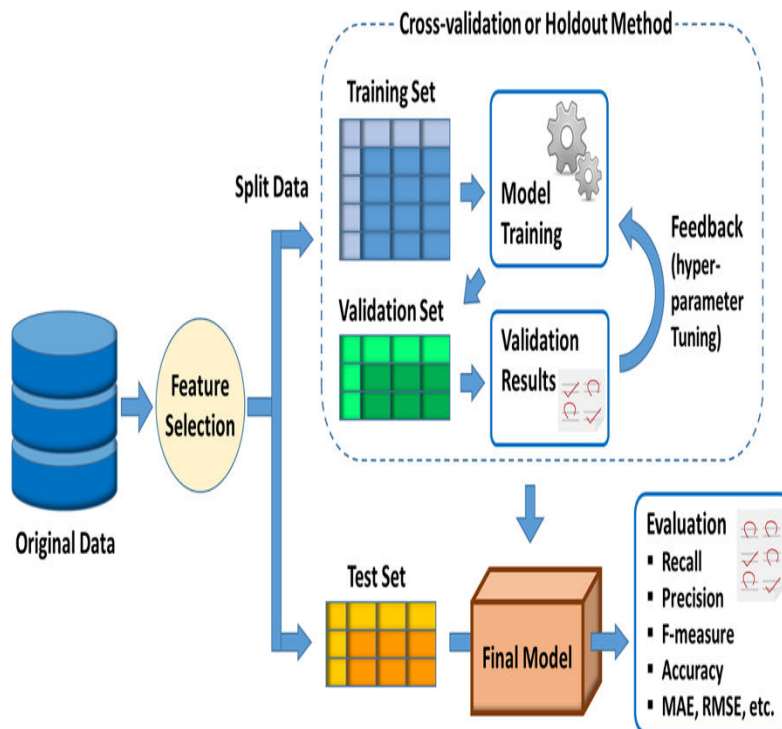
Random Forest Classifier



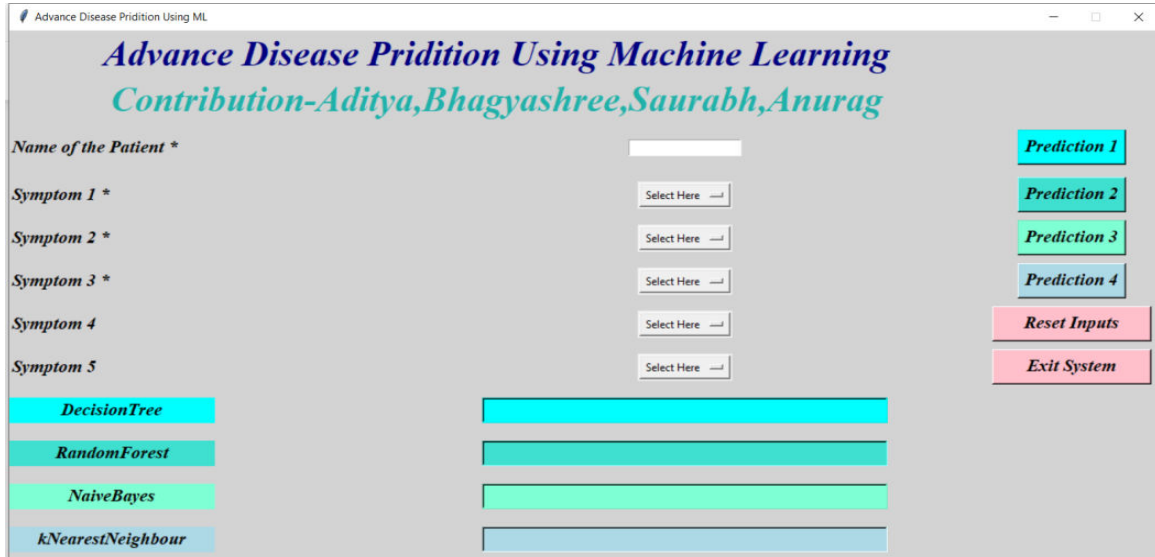
K Nearest Neighbour- It is a supervised learning algorithm. It is a basic yet essential algorithm. It finds extensive use in pattern finding and data mining. It works by finding a pattern in data which links data to results and it improves upon the pattern recognition with every iteration. We have used K Nearest Neighbour to classify our dataset and achieved ~92% accuracy.



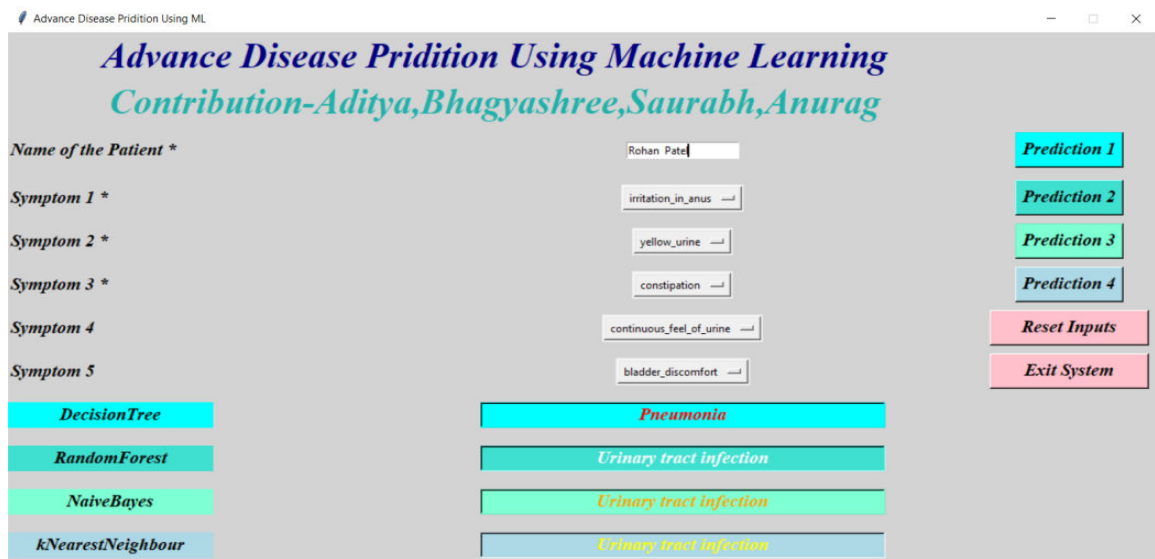
Naïve Bayes algorithm- It is a family of algorithms based on naïve bayes theorem. They share a common principle that is every pair of prediction is independent of each other. It also makes an assumption that features make an independent and equal contribution to the prediction. In our project we have used naïve bayes algorithm to gain a ~95% accurate prediction.



V. RESULTS



GUI Interface



Predicted Output

VI. CONCLUSION

We set out to create a system which can predict disease on the basis of symptoms given to it. Such a system can decrease the rush at OPDs of hospitals and reduce the workload on medical staff. We will use 4 different algorithm to do so. On an average we achieved accuracy of ~94%. Such a system can be largely reliable to do the job. Creating this system we also added a way to store the data entered by the user in the database which can be used in future to help in creating better version of such system. Our system also has an easy to use interface. It also has various visual representation of data collected and results achieved.



VII. FUTURE SCOPE

This project has not implemented recommendation of medications to the user. So, medication recommendation can be implemented in the project. History about the disease for a user can be kept as a log and recommendation can be implemented for medications.

REFERENCES

- [1] <http://people.dbmi.columbia.edu/~friedma/Projects/DiseaseSymptomKB/index.htm>
- [2] <http://www.livescience.com/36519-diseasesconditions-symptoms-treatments.html> Live Science
- [3] http://www.saedsayad.com/naive_bayesian.htm
- [4] DR Mahboob Khan (PhD) “Smart Health Prediction using Data Mining” , International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)
- [5] Daniel Lowd ,PedroDomingos “Naïve Bayes Model For Probability Distribution” , Department of Computer Science and Engineering, University of Washington, Seattle, WA 98195-2350, USA
- [6] Sujatha, Sumathy, Anitha Nithya “A Survey of Health Care Prediction Using Data Mining”, International Journal of Innovative Research in Science, Engineering and Technology Vol. 5, Issue 8, August 2016



INNO  **SPACE**
SJIF Scientific Journal Impact Factor
Impact Factor: 7.542



ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

 **9940 572 462**  **6381 907 438**  **ijircce@gmail.com**



www.ijircce.com

Scan to save the contact details