Journal of Theoretical and Applied Information Technology
30th April 2024. Vol.102. No 8

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

EMPIRICAL INVESTIGATIONS TO DISEASE PREDICTION USING MACHINELEARNING AND WEB INTEGRATION

[1]MR ANIL KUMAR PALLIKONDA, [2]R ABHINAYA, [3]PVRS PADMARAJU, [4]DR CHANDANAPALLI SURESH, [5]DR DNVSLS INDIRA, [6]SUDHAKAR ATCHALA,

[1]Dept. of CSE, PVP Siddhartha Institute of Technology, Vijayawada, Andhra Pradesh, India.
[2]Department of CSE , Seshadrirao Gudlavalleru College of Engineering, Gudlavalleru, AP, India
[3]Departmet of Mechanical Engineering, SRKR College of engineering, Bhimavaram, AP, India
[45]Department of IT , Seshadrirao Gudlavalleru College of Engineering, Gudlavalleru, AP, India
[6]Dept. of CSE, Lakkireddy BaliReddy College of Engineering, Mylavaram, Andhra Pradesh
Email: anilkumar.pallikonda@gmail.com

ABSTRACT

Health has become the utmost priority of people's lives after the whole pandemic situation. Many new emerging technologies are striving for better human livelihood. In technology, particularly Machine Learning has a great potential in making Human lifestyle better and much simpler. Machine learning has extended its arms in Marketing Banking, Smart life, Production and Manufacturing and widely in the health care sector.ML has a great impact on the healthcare industry ML is widely used for prediction of Cancer, prediction and identifying the damage in kidneys, heart disease predictions etc. By integrating the complex algorithms with real-time medical training datasets can produce high accurate results. Hospitals and doctors have been extremely busy during the COVID period. There was a time to avoid visiting hospitals during these times unless it was absolutely necessary. This application takes the common symptoms of the patient and tries to give possible predictions as a primary disease. The application is built by using the most famous algorithms in the ML domain like Decision Tree classification algorithm, Random Forest classification algorithm, Naïve bayes classifier algorithm. This application shows the mechanism of three algorithms for disease prediction

Keywords: *Random Forest, Naïve Bayes Classifier, Disease Prediction*

1. INTRODUCTION

In today's world, Technology is playing a key role in the Healthcare industry , with evolving technology and changing times things have not been the same and in fact have become much easier. Disease Prediction using Machine learning is an attempt to do that. This Disease Prediction system basically uses the information that is provided by the users and comes up with a prediction. In broader terms, if any patient is feeling ill but they feel it's not very serious momentarily , but they still want to know what they could be suffering from , they can just enter the symptoms and the Disease prediction system provides them with the possible disease they might be suffering from. This way, the patient does not have to go and consult a doctor even for small things and this saves a lot of time too and to most extent gives accurate results as well. This system can be built solely using Machine Learning algorithms using python and in order to make it more user friendly we have created a website where the patient can select the symptoms and get the predictions just by a click. This was achieved by integrating the Machine learning algorithms with Flask

2.LITERATURE REVIEW

S Radhika, S Ramiya Shree, V Rukhmani Divyadharsini, and A Ranjitha proposed and in which they considered a Decision tree algorithm for detection of diseases. In this they compared different algorithms for accuracy and plotted a graph and found out that Decision tree has shown better accuracy in comparison, which led us to choosing Decision Tree algorithm for our model.

Md. Zahangir Alama, M. Saifur Rahman, M. Sohel Rahman, proposed on the prediction of the disease using random forest led us to using this algorithm. In this paper many datasets were given, and each dataset was classified based on diseases, so the accuracy was predicted.

Mostafa Langarizadeh and Fateme Moghbeli proposed and in NCBI talked about applying

Journal of Theoretical and Applied Information Technology
30th April 2024. Vol.102. No 8

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Naive Bayes Algorithm, a ROC curve has been plotted for the accuracy comparison the help of the symptoms given by the patient.
Sneha Grampurohit and Chetan Sagarnal proposed disease prediction using machine learning algorithms has considered basically three algorithms for prediction: Naïve bayes classifier, Random forest, Decision tree algorithm , all of which we used in our project. In this sample datasets from a standard resource have been considered and have been implemented on the system we got a clear idea with the help of that.

3.STAMENT OF THE PROBLEM

In recent times, with the emergence of Covid and various other reasons, the Healthcare industry has been one of the busiest industries globally. Hospitals and Doctors have been very busy with everything that's going on and giving consults to easily identifiable diseases is a waste of both time and resources . Also, it is not very advisable for people to go to hospitals unless it's absolutely necessary as it may lead to unnecessary problems. So in order to overcome this problem , it would be really helpful if there was a way for the users or patients to know what they are suffering from just by providing some information about their symptoms. That is when we came up with the idea of a Disease prediction model through which the patient can just enter their symptoms and get predictions of what the disease might be. This could help a lot of people of all age groups. The objective of this project is to use a dataset directly retrieved from Kaggle and this is based on past hospital datasets, therefore on comparison the accuracy must increase. The main objective is to use the Machine learning algorithms and build models such that we get accurate disease predictions according to the symptom.

Block Diagram of the proposed method can be seen as below

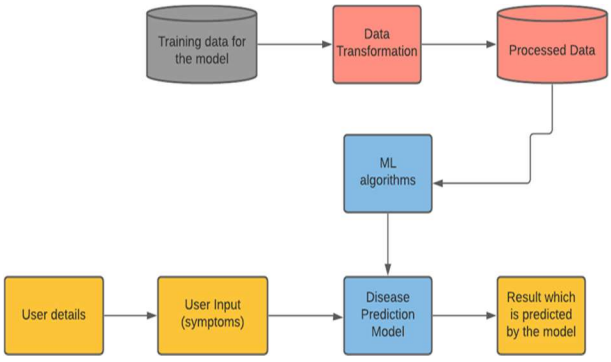


Figure 1: Block Diagram Of Proposed Work

4. PROBLEM METHODLOGY AND SOLUTION

The Disease Prediction model using Machine Learning and Web Integration is implemented completely using Python. Flask which is used in web integration is also written using python. Machine learning gives computers the ability to learn. It gives the computers the ability to learn without being programmed explicitly

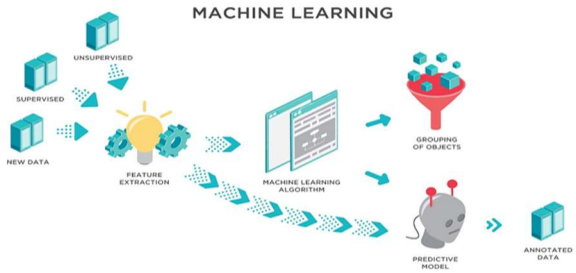


Figure 2 Overview Of Machine Learning

Machine Learning Algorithms used for building this model are:

4.1Decision Tree Algorithm:

Decision tree is a tree structure which resembles a flowchart.
It is a non- parametric approach which is used for building classification models and finding an optimal decision tree is a complete problem in the NP approach.
The methods used for the development of decision trees are less expensive which makes it easier to develop models even when the training set size is large.
Entropy formula for Decision tree algorithm is as below:

$$E(S) = -p_{(+)}\log p_{(+)} - p_{(-)}\log p_{(-)}$$

And Information gain = E(Y)-E(Y/X)

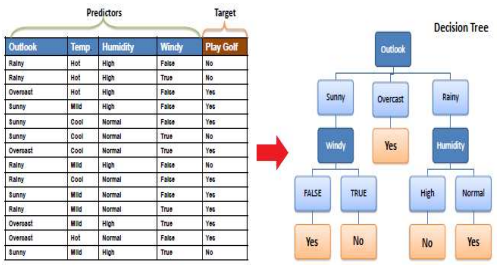


Figure 3 Example Of Decision Tree

Journal of Theoretical and Applied Information Technology
30th April 2024. Vol.102. No 8

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

4.2 Random Forest:

Random Forest is a classifier which consists of multiple decision trees.

These decision trees operate together as an ensemble.

Each tree gives a prediction and the one with the most votes becomes the final prediction.

The prerequisites for random forest to perform well are:

There needs to be some actual signal in our features so that models built using those features do better than random guessing.

The predictions made by each of the trees must have low correlations with each other

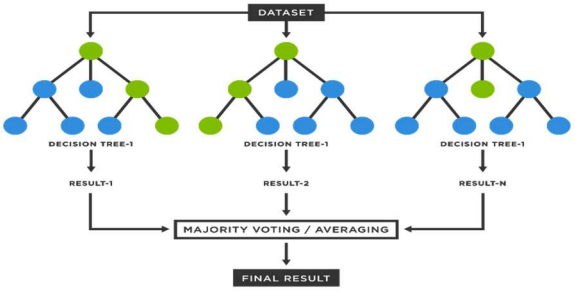


Figure 4 Overview Of Random Forest

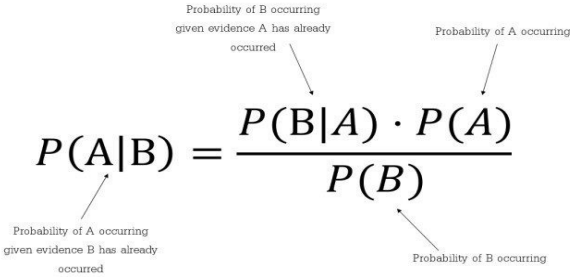
4.3 Naive-Bayes Classifier:

A group of classification algorithms built on the Bayes' Theorem are known as naive Bayes classifiers. It is a family of algorithms rather than a single method, and they are all based on the idea that every pair of features being classified is independent of the other.

A straightforward mathematical procedure for computing conditional probabilities is known as Bayes' Theorem.

The probability of an event happening given that another event has already occurred (via assumption, presumption, statement, or evidence) is known as conditional probability.

The equation is:



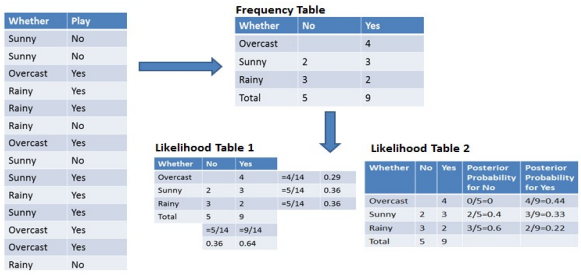


Figure 5 Example Of Naïve Bayes Classifier

4.4 Flask:

A Python package called Flask serves as a web framework that makes it simple to create web apps. Its core is compact and simple to extend; it's a micro framework.

It does have a lot of great features, including a template engine and url routing. It is a web app framework for WSGI.

Provides libraries to build lightweight web applications in python.

Figure 6. Flask Logo

5. Testing

White Box Test cases:

Check if the environment is set correctly.

Check if all the codes are integrated and functional.

If all the algorithms are functioning properly or not.

Black Box Test cases:

And in the process of testing, we also tested the website and its working quality. After compiling the app.py it compiled successfully and rendered the home.html on web browser in local host port http://127.0.0.1:5000/

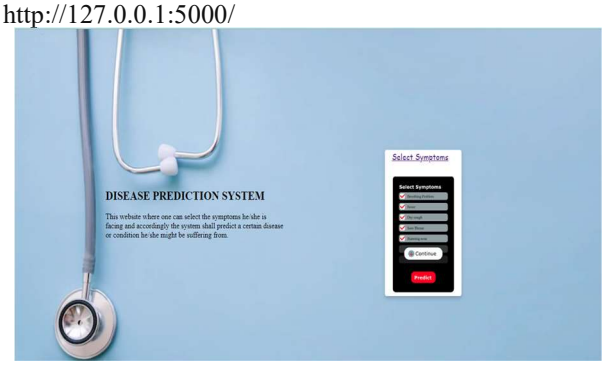


Figure 7. Home Page

On selecting the select symptoms it redirects to the index.html page where you can select symptoms

Journal of Theoretical and Applied Information Technology
30th April 2024. Vol.102. No 8

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

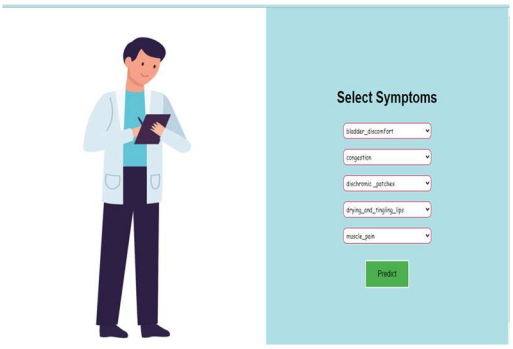


Figure 8 Page To Select Symptoms

Where we have selected 5 symptoms as the input by using dropdown and click select on Predict button

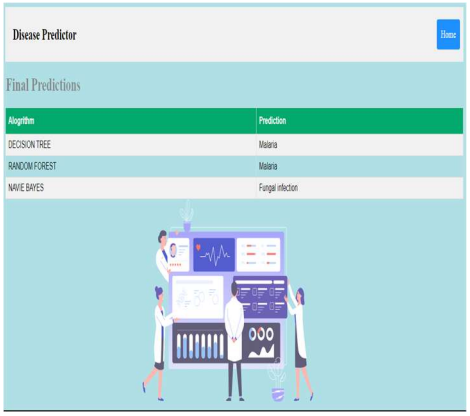


Figure 9 Final Predictions

Here are our final Results, Decision Tree is predicting Malaria ,Random Forest is predicting Malaria, Naive Bayes is Predicting Fungal Infection.After that by clicking home we will redirect to the home page.

Test Cases

As Testing case-1 i gave symptoms “mild_fever”, “neck_pain”, “runny_nose”, “cramps, “abdominal_pain”, as input through dropdown boxes

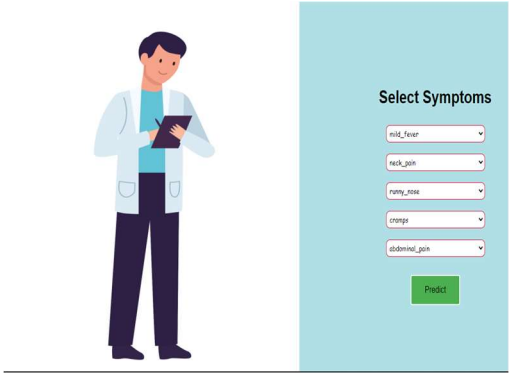


Figure 9 Page To Select Symptoms

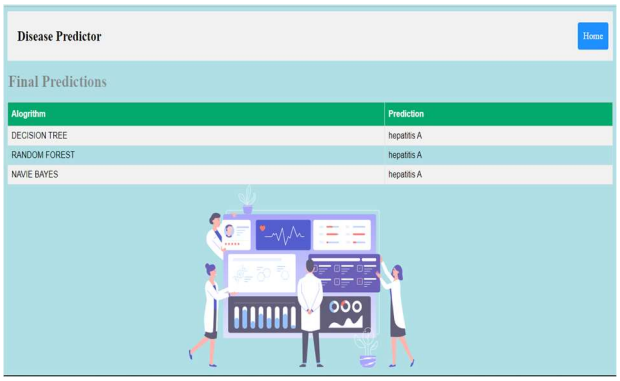


Figure 10 Page To Select Symptoms

All the Algorithms show hepatitis A as the prediction in this case and here we assume that the patient is suffering from hepatitis A disease and the model is working Flawlessly with these inputs.

8.RESULTS AND DISCUSSIONS

To evaluate the models, we have a wide range of evaluation techniques in the machine learning domain. Here in this project, we have used the accuracy metric to evaluate all the 3 models. By calculating the accuracy, we get the following results:

Figure11. Results About Disease

9. CONCLUSION AND FURTHER SCOPE:

So, we can conclude that on the basis of the symptoms entered by the user in the website we get three predictions from the three algorithms and we get an overall accuracy of 95.12 percent. Further scope of this project is that its features can be extended by giving the diagnosis for the prediction received, showing the nearby pharmacy etc. Also nearest hospitals or clinics which treat the predicted disease can be suggested.

REFERENCES

[1] Aldahiri, Amani, Bashair Alrashed, and Walayat Hussain. "Trends in using iot with machine learning in the health prediction system." Forecasting 3.1 (2021): 181-206.

[2] Ahmad, Muhammad Aurangzeb, Carly Eckert, and Ankur Teredesai. "Interpretable machine learning in healthcare." Proceedings of the 2018 ACM international conference on bioinformatics, computational biology, and health informatics. 2018.
[3] Radhika, S., et al. "Symptoms Based Disease Prediction Using Decision Tree and Electronic Health Record Analysis." European Journal of Molecular Clinical Medicine, vol. 7, no. 4, 2020, pp. 2060-2067.
[4] Tu, P-L., and J-Y. Chung. "A new decision-tree classification algorithm for machine learning." TAI'92- Proceedings Fourth International Conference on Tools with Artificial Intelligence. IEEE Computer Society,1992.
[5] Karthika, S., and N. Sairam. "A Naïve Bayesian classifier for educational qualification." Indian Journal of Science and Technology 8.16 (2015): 1-5.