ZKG INTERNATIONAL

ISSN: 2366-1313

A VARIOUS SYMPTOMS BASED DISEASE PREDICTION USING MACHINE LEARNING

SK. AHAMAD SHARIF[1], CH. GOPI[2], B. PRASANNA LAKSHMI[3], B.VENKATA SAI RAHUL[4], B. SONY BHARGAVI[5].

[1] Assistant Professor, CSE,Chalapathi Institute of Technology,Guntur, India
[2]UG Student,CSE,Chalapathi Institute of Technology,Guntur, India
[3]UG Student,CSE,Chalapathi Institute of Technology,Guntur, India
[4]UG Student,CSE,Chalapathi Institute of Technology,Guntur, India
[5]UG Student,CSE,Chalapathi Institute of Technology,Guntur, India

ABSTRACT: In recent times, researchers have conducted many experiments to design various automated analysis models using different supervised machine learning models. Machine learning technology has been proven to be beneficial in the medical field so that health care issues can be resolved effortlessly and efficiently. Prediction of a disease at an early stage might help diagnose those diseases, which may control the death rate caused by those diseases. These days, many virtual prediction models are available in the market at low cost. The existing systems made us use an algorithm that would predict a disease that may or may not be accurate. The maximum accuracy of the existing system's range is not even more than 88%. The algorithms used in the prediction systems consisted of various machine learning algorithms. In our project, "Disease Prediction Based on Symptoms Using Machine Learning," it is possible to predict a disease with higher accuracy. So, the user does not need to traverse many platforms to predict a disease. It will help to reduce the time, cost, and resources of predicting a disease at earlier stages, so as to prevent the extremities of it, and thus, there is a higher chance of reducing the mortality rate.

1. INTRODUCTION

Healthcare has become the essential element in human lifestyle and economy. There is a huge change in the world as we know it when compared to the present living conditions compared to the world that existed few years back. Everything has turned unpleasant and divergent. In this case, where the entirety has grown to become digital or let us just say virtual, the doctors and nurses are giving their maximum effort to keep people's lives and their health stable even when it comes to risking their very own in the process. Even now some parts of the world there are still some distant villages, remote places which lack the required health facilities. Machines have started to gain popularity in the 21st century and dependency on humans has been decreased, without any human mistakes they could perform duties with greater efficiency and a steady degree of accuracy.

A disease predictor is nothing but a virtual doctor, which can predict the disease of any affected person without any human errors and it happen to do so without using much resources. During the early stages the

ISSN: 2366-1313

disease prediction system focused more on input of blood sample reports. Whereas the proposed system 'Disease Prediction Based on Symptoms' Model helps to predict the disease of the patient based on the input of his/her symptoms. Depending on the disease predicted by the proposed system a specialized doctor will be assigned to that patient.The following algorithms used in developing the Disease Prediction Based on Symptoms Model include:

1**. Decision Tree:** it belongs to the family of supervised learning algorithms. Unlike other the decision tree algorithm can be used for solving regression and classification problems. It is a tree-structured classifier, where internal nodes represent the attributes or features of a dataset and the branches represent the decision rules and each leaf node represents the outcome. It is used to create a training model that can be used for predict the class or value of the target variable by learning simple decision rules deduced from training data.

2. Random Forest: It is a popular machine learning algorithm; it belongs to the supervised learning technique. It can be used to solve both Classification and Regression problems in the Machine Learning. It is based on the concept of ensemble learning, which is a process of combing multiple classifiers to solve a many complex problems and to improve the performance of the model.

3. Naïve Bayes algorithm: it is a supervised learning algorithm; it is based on Bayes theorem and used for solving classification problems. it is highly scalable and is one of the simplest and most effective Classification algorithms, which helps in building the fast machine learning models that can make quick predictions.

Naive Bayes algorithm models are known under various names like simple Bayes and independence Bayes.

4. Gradient Boosting Algorithm: it is a machine learning technique used in both regression and classification. It gives a best prediction model in the form of an group of weak prediction models. The gradient boosting technique is a type of machine learning boosting. It relies mostly on the intuition that the best possible next model, when combined with prior models, minimizing the overall prediction error.

2. LITERATURE SURVEY

Naïve bayes variants in classification learning. By Khadija Mohammad Al-Aidaroos, Azuraliza Abu Bakar, and Zalinda Othman.

These researchers have conducted a study to determine the best medical diagnosis mining technique. They compared Naïve Bayes with five other classifiers, including Linear Regression classifier, K-Star classifier(K*), Decision Tree (DT), Random Forest, and rule-based algorithm (ZeroR). The researchers evaluated the efficiency of all algorithms using 15 real-world medical problems from the UCI machine learning repository. The results indicated that Naïve Bayes outperformed the other algorithms in 8 of the 15 data sets, demonstrating its superior predictive accuracy.

Predicting individual disease risk based on medical history.by Darcy A. Davis, Nitesh Vijay Chawla, Nicholas Blum, Nicholas Christakis, and Albert-Laszlo Baribas.

These researchers have discovered that treating chronic illnesses globally is neither time nor cost-effective. Therefore, they conducted a study to predict potential disease risks using CARE, a system that predicts possible disease risks using only a patient's medical history and ICD-9-CM codes. CARE incorporates collective filtering approaches with clustering to predict each patient's greatest disease risks. The authors also defined an iterative version

ISSN: 2366-1313

called ICARE, which integrates ensemble principles to improve efficiency. These systems can predict a wide range of medical conditions accurately and efficiently without advanced knowledge. ICARE's exceptional potential risk coverage results in more accurate early warnings for numerous diseases years ahead of time. The CARE system can investigate a wider range of disease backgrounds, raise unconsidered questions, and facilitate discussions regarding early detection and prevention.

3. EXISTING SYSTEM: -

Machine can predict diseases but they are capable of predicting only limited types of the diseases. It fails to predict all possible conditions of the people. Existing system handles only structured data. The prediction system is broad and ambiguous.

In current past, countless disease estimate classifications have been advanced and in procedure. The standing organizations arrange a blend of machine learning algorithms which are judiciously exact in envisaging diseases. However the restraint with the prevailing systems is speckled. First, the prevailing systems are dearer only rich people could pay for to such calculation systems. And also, when it comes to folks, it becomes even higher. Second, the guess systems are non-specific and indefinite so far. So that, a machine can envisage positive disease but cannot expect the sub types of the diseases and diseases caused by the existence of one bug. For occurrence, if a group of people are foreseen with Diabetes, doubtless some of them might have complex risk for Heart viruses due to the actuality of Diabetes. The remaining schemes fail to foretell all possible surroundings of the tolerant.

4. PROPOSED SYSTEM

Our application will be at affordable cost. Decision Tree Machine Learning Algorithm predicts Diseases as well as all sub diseases. Algorithm is implemented to increase operational efficiency. It reduces Query retrieval time.

Accuracy is improved using Machine Learning algorithm. The proposed system begins with the thought that was not executed by the ancestors. It gadget Decision Tree machine learning procedure for calculating diseases as well as calculating all the other thinkable diseases. Its member Map Reduce algorithm for subdividing the data such that a request would be scrutinized only in the explicit partition, which will increase effective proficiency but cut query rescue time. In tally to that, it provides definite rations for specific clients to pattern his/her condition. Thus making our presentation broadly open by all at cheap cost.

5. SYSTEM ARCHITECTURE

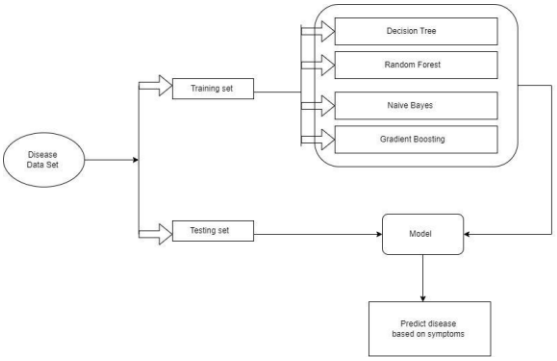


Fig: Architecture Disease Data Set

6. IMPLEMENTATION

Dataset and Model Description

The system utilizes a structured dataset containing the name of diseases and their corresponding symptoms. As this system is based on supervised machine learning algorithms, the dataset includes a label of 0 or 1. The dataset is then divided into a training dataset and a testing dataset. The model is trained using the training dataset,

ISSN: 2366-1313

where all algorithms are applied to it, and the machine learning model is trained. To test the accuracy of the model, the testing dataset is provided to the trained model.

Dataset of Hospital

The format of the hospital data will be structured. The dataset utilized in this project consists of real-life data regarding patients' symptoms. The structural data contains information about symptoms of patients. Any dataset is binary, with a value of 0 indicating the symptom has an impact on the disease and a value of 1 indicating that it does not impact the disease.

Evaluation Method

To evaluate performance in the experiment, we initially identify TP, TN, FP, and FN as True Positive (the number of results predicted correctly as required), True Negative (the number of results predicted correctly as not required), False Positive (the number of results predicted incorrectly as required), and False Negative (the number of results predicted incorrectly as not required), respectively.

We can calculate four metrics, including recall, precision, accuracy, and F1 measures using the following formulas:

Accuracy-:

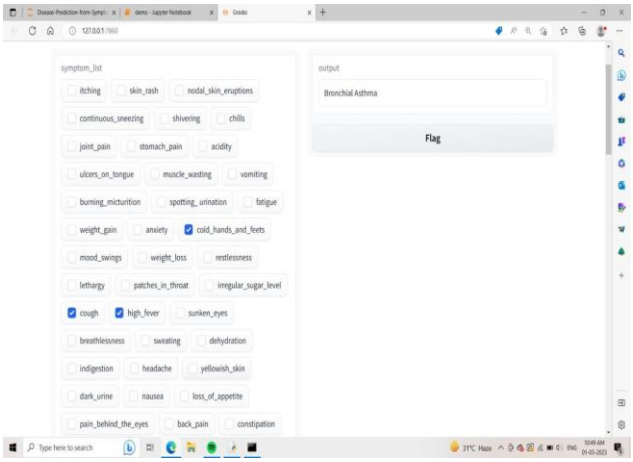
$$\frac{TruePositive+TrueNegative}{TruePositive+TrueNegative+FalsePositive+FalseNegative}$$

$$Precision = \frac{TruePositive}{TruePositive+FalsePositive}$$

$$Recall = \frac{TruePositive}{TruePositive+FalseNegative}$$

$$F1\text{-}Measure = \frac{2\times precision\times recall}{precision+recall}$$

7. SCREEN SHORT



8. CONCLUSION

Making predictions from data is a strong use of machine learning. But it\'s crucial to keep in mind that machine learning is only as effective as the data used to train the algorithms. The website has been created in such a way that using it will be simple and easy for users. The project successfully implemented a website that could predict a disease with a 97% accuracy rate after comparing the four algorithms. Along with disease prediction, the website includes an about page, consultancy page and a feedback page where users can provide valuable feedback. In terms of future work, we intend to store the data of the users and use that information in the existing dataset and work on increasing the accuracy rate as well as include a greater number of diseases which the model can predict, generation of report and include services like virtual doctor appointments and online medicine shopping.

REFERENCE

[1]Maniruzzaman, M., Rahman, M., Ahammed, B. and Abedin, M., 2020. Classification and prediction of diabetes disease using machine learning paradigm. Health information scienceand systems, 8(1), pp.1-14.

[2] Chen, M., Hao, Y., Hwang, K., Wang, L. and Wang, L., 2017. Disease prediction by

ISSN: 2366-1313

machine learning over big data from healthcare communities. Ieee Access, 5, pp.8869-8879.
[3] Haq, A.U., Li, J.P., Memon, M.H., Nazir, S. and Sun, R., 2018. A hybrid intelligent systemframework for the prediction of heart disease using machine learning algorithms. Mobile Information Systems, 2018.
[4] Keniya, R., Khakharia, A., Shah, V., Gada, V., Manjalkar, R., Thaker, T., Warang, M. and Mehendale, N., 2020. Disease prediction from various symptoms using machine learning.
Available at SSRN 3661426.
[5] Dahiwade, D., Patle, G. and Meshram, E., 2019, March. Designing disease prediction model using machine learning approach. In 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC) (pp. 1211-1215). IEEE.
[6] Kaushik, K., Kapoor, D., Varadharajan, V. and Nallusamy, R., 2014. Disease management: clustering–based disease prediction. International Journal of Collaborative Enterprise, 4(1-2),pp.69-82.