International Journal of Combined Research & Development (IJCRD)
eISSN:2321-225X;pISSN:2321-2241 Volume: 11; Issue: 6; June -2022

Disease Prediction for Various Symptoms using Machine Learning

Pooja M M
PG Student, MCA Dept.
JNNCE
Shimoga
poojamm777@gmail.com

Prashant Ankalkoti
Asst. Prof, Dept of MCA
JNNCE
Shimoga
prashantsa@jnnce.ac.in

Abstract : The Data mining Techniques play an vital role in developing the several real world applications such as Industries ,Healthcare ,E-commerece etc. All this applications utilizes the above techniques in machine learning. In the Healthcare industry, a healthcare provider are looking forward for robust and reliable healthcare system and the aim of developing the robust system is to diagnose the health related issue in the initial stage.

Keywords : Data Mining, Random forest classifier, K-Nearest Neighbor(KNN), Machine learning

.

1. INTRODUCTION

In the modern lifestyle the technology and the research field are rapidly developed in the society. Technology is in healthcare play an important role during the pendamic situation. In the healthcare industry large number of data is generated and it is difficult to analyze the large amount of data, So the Data mining techniques is used. This techniques will helps in collecting the useful information from the sample data and also helps to health officials to improve the healthcare service and can provide a early treatment for health issue.In real time, the data is collected from patient's based on their symptoms and data will be trained using ML algorithm through this algorithms the disease will be predicted. Machine Learning is a trendy innovation that can increase the several industry and skilled process and it present everywhere in the world so even unknowingly many people's are using this daily.

2. LITERATURE SURVEY

Dhiraj Dahiwade et al., [1] has published the paper which has made the comparison between the Convolutional Neural Network(CNN) and K- Nearest Nieghbor(KNN). Here CNN will give more accuracy and requires less time for classification compared to KNN.

Nishant Yede et al.,[5] has developed a clinical system and it will helps to patient's to get a proper treatment. Clinical system was developed by using an several ML algorithms are used for predicting the initial stage observation of disease. It gives 94.8% of accuracy Pahulpreet Singh Kohli et al.,[7] has implemented the Support Vector Machine, Logistic Regression, Adaptive Boosting(AdaBoost classifier , Decision Tree, Random forest Algoritms. This Algorithms achieved the accuracy of 85. 71%, 87.1%, 98.57% for Predicting the early detection of disease.

Priyanka J. Panchal et al.,[8] has published the paper,it used a Random Forest[RF] Algorithm for forecast the disease for various indicator given by the user and by using an algorithm will predict the disease that will helps to the doctors.

Sneha Grampurohit et al.,[10] has published the paper which has implemented Decision Tree[DT] classifier, Random Forest[RF] classifier , Naive Bayes[NB] Classifier to predict disease. This paper made the comparative study between the above algorithms with the accuracy of 94% .

3. PROPOSED METHODOLOGY

The proposed methodology is mainly focused on predicting the diseases. It provide two classification method that is RF and K-NN .Based on the indictor selected the disease will be predicted.

3.1 Algorithm

Step 1: Read the values from the user

Step 2: Store the patient name and 3 symptoms in variables
if variable is equal to null then goto step1
else goto step 2

Step 3: Click the following classifiers using which you want to predict
- Random forest
- K-Nearest Neighbor

Step 4: Predict the disease based on the symptoms using selected classifier

Step 5: Display the accuracy of the classifier.

3.2 Block Diagram

Block diagram is a diagrammatic representation of the proposed work. It will help the user to easily understand the flow of the proposed work by showing brief of each step of the work. Figure 1 displays the pictorial representation of the proposed work.

The training and testing databanks used in the proposed work is taken from kaggle.com. This dataset is manipulated according to

the need of methodology , Then the dataset will be trained by applying the following classifier that is KNN and Random forest classifier.

The system will ask the users to enter their information ,select their symptoms and the disease will be predicted based on ML algorithms, Later the Accuracy of the Algorithm is also predicted.

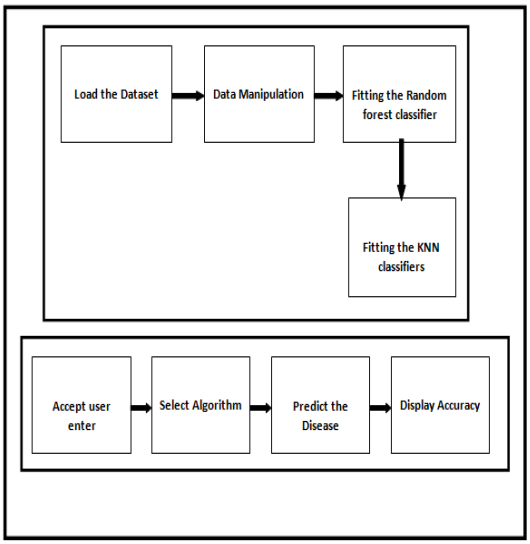

Figure 1: Block diagram of the project

4. RESULT ANALYSIS

Result Analysis is a section where it will analyzes the result that obtained through the proposed work and using machine learning algorithm the disease will be predicted and it will display the result in gui, calculates the accuracy of an algorithm and plotiing the graph for the accuracy. The result of proposed work as given in the Figures

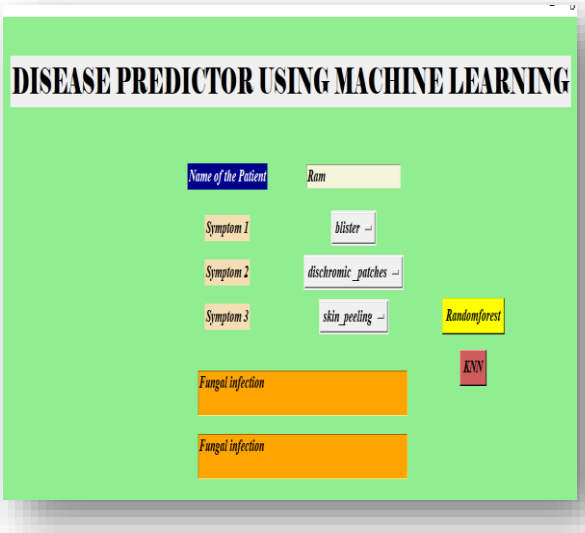

Figure 2 Result of an GUI

In this Figure 2 the gui contains 1 text box where user need to enter the patient name and 3 drop box is present in that user need to select symptoms .After selecting the symptoms they have to select the algorithm based on which the disease will be predicted and displayed in another text box to identify the system .

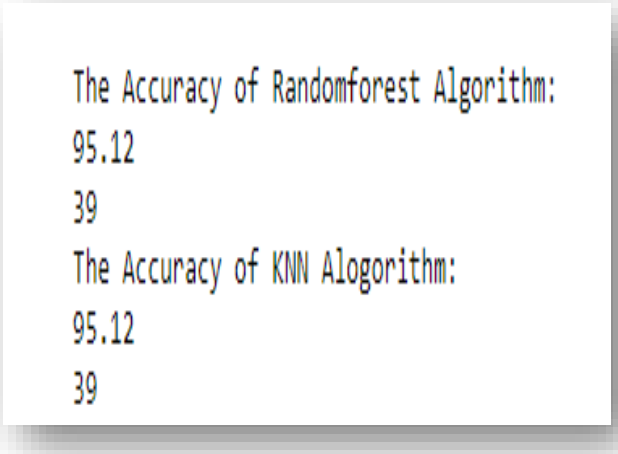

Figure 3 Accuracy score and No of disease predicted is true

As observed in Figure 3 it clearly visible that both the algorithm that is K-Nearest Neighbor and Random Forest has same accuracy of 95.12%. It is also displaying how many number of Samples (disease) has been predicted true. Both the algorithm predicting 39 samples as true.

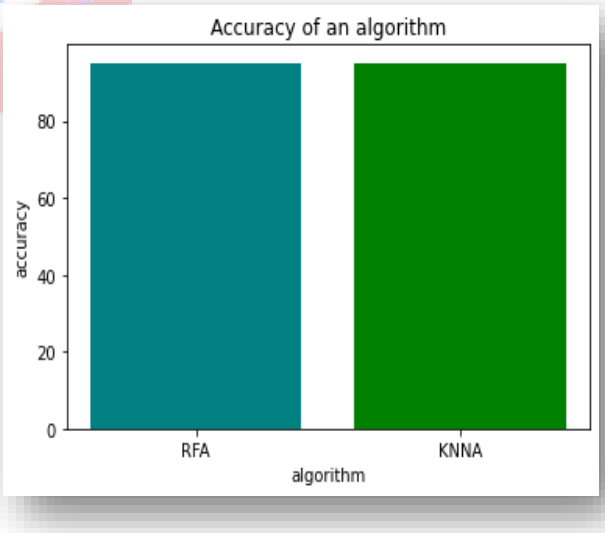

Figure 4 Accuracy Graph

In the Figure 4 It representation of bar graph. It is plotted on accuracy versus algorithm. The both algorithm is displaying an accuracy of 95.12%.

5. CONCLUSION

The proposed work is discussing about the application of Machine

learning model to predict the illness based on indicators using Random forest [RF] and K-Nearest Neighbor algorithms. Various symptoms are provided in the drop down menu, from listed menu user selects any three of them and then using 2 different algorithms the disease will be predicted. The accuracy score of random forest classifier is 95.12% and K-Neighbor classifier is 95.12%.

6. REFERENCES

[1] D. Dahiwade, G. Patle and E. Meshram, "Designing Disease Prediction Model Using Machine Learning Approach," 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), 2019, pp. 1211-1215, doi: 10.1109/ICCMC.2019.8819782.

[2] G. Zhang, "A Modified SVM Classifier Based on RS in Medical Disease Prediction," 2009 Second International Symposium on Computational Intelligence and Design, 2009, pp. 144-147, doi: 10.1109/ISCID.2009.43.

[3] H. Q. Yu, "Experimental Disease Prediction Research on Combining Natural Language Processing and Machine Learning," 2019 IEEE 7th International Conference on Computer Science and Network Technology (ICCSNT), 2019, pp. 145-150, doi: 10.1109/ICCSNT47585.2019.8962507.

[4] M. Krishnamoorthy, M. S. A. Hameed, T. Kopinski and A. Schwung, "Disease Prediction Based on Individual's Medical History Using CNN," 2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA), 2021, pp. 89-94, doi: 10.1109/ICMLA52953.2021.00022.

[5] Nishant Yede, Ritik Koul, Chetan Harde ,Kumar Gaurav "General disease prediction based on symptoms provided by patient" Open Access International Journal of Science & Engineering. Volume 6 June 2021 DOI 10.51397/OAIJSE06.2021.0032

[6] P. Jha, T. Biswas, U. Sagar and K. Ahuja, "Prediction with ML paradigm in Healthcare System," 2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC), 2021, pp. 1334-1342, doi: 10.1109/ICESC51422.2021.9532752.

[7] P. S. Kohli and S. Arora, "Application of Machine Learning in Disease Prediction," 2018 4th International Conference on Computing Communication and Automation (ICCCA), 2018, pp. 1-4, doi: 10.1109/CCAA.2018.8777449.

[8]] Priyanka J. Panchal , Shaeezah A. Mhaskar , Tejal S. Ziman "Disease Prediction using Machine Learning" Iconic Research And Engineering Journals Volume 3 Issue 10 2020

[9] R. C. Mallela, R. L. Bhavani and B. Ankayarkanni, "Disease Prediction Using Machine Learning Techniques," 2021 5th International Conference on Trends in Electronics and Informatics (ICOEI), 2021, pp. 962-966, doi: 10.1109/ICOEI51242.2021.9453078.

[10] S. Grampurohit and C. Sagarnal, "Disease Prediction using Machine Learning Algorithms," 2020 International Conference for Emerging Technology (INCET), 2020, pp. 1-7, doi: 10.1109/INCET49848.2020.9154130.