Disease Prediction Using Machine Learning

Asmita Hanmant Halde, Priti Suresh Bilambe, Pratiksha Uamesh Bachate,
Aishwarya Narayan Bolabattin
Department of Computer Science
Shree Siddheshwar Women's Polytechnic College of Diploma, Solapur, Maharashtra, India

Abstract: *The project implements 3 linear models and one deep learning model: Naïve Bayes, Support Vector Machine, K-Nearest Neighbors network to investigate their performance on diabetes and heart disease datasets obtained from the UCI data repository. In addition to the comparison of the algorithms, each algorithm has been integrated into a prediction engine and exposed over an API. The project also includes a web platform to facilitate collaboration among researchers and doctors. As the results show, our prediction engine is capable of recognizing the presence of the disease and also predict it accurately. Performance improvements could also be achieved by using complex deep learning methods are Disease Prediction using Machine Learning is the system that is used to predict the diseases from the symptoms which are given by the patients or any user. The system processes the symptoms provided by the user as input and gives the output as the probability of the disease. an increase in biomedical and healthcare data, accurate analysis of medical data benefits early disease detection and patient care. By using linear regression and decision tree we are predicting diseases like Diabetes, Malaria, Jaundice, Dengue, and Tuberculosis.*

Keywords: Tuberculosis

I. INTRODUCTION

Today's Healthcare organizations produce and collect large volumes of information on a daily basis. Looking at the need of patient's, the proposed system will be implemented to save the time of patients'. Human disease is the main reason for human death.

The proposed system able to predict disease by adding symptoms related to patients' condition and behavior, which is accomplished by data analysis. Raw data from healthcare are heterogeneous they need to collect and store in organized forms, this data used for early detection of human disease. In the proposed system we enter symptoms, as per patients' condition, the system analyzes symptoms given by the patient and predicts diseases.

This proposed system will not only predict the diseases but also recommend the appropriate doctors based on a particular disease. The list of doctors' datasets will be used for both the symptoms checking and prediction of diseases. This proposed system will predict diseases like polio, dengue, lung disease, blood cancer. This system is user-friendly. Data mining also is known as Knowledge Discovery in databases refers to the nontrivial extraction of implicit, previously unknown and potentially useful information from data stored in databases. KDD is an iterative process where evaluation measures can be enhanced, mining can be refined, new data can be integrated and transformed in order to get different and more appropriate results. While data mining and knowledge discovery in databases are frequently treated as synonyms, data mining actually part of the Knowledge Discovery process. The following figure shows data mining as a step in the knowledge

What is Disease Prediction?

It is way to recognize the patient health by applying data mining and machine learning techniques on patient treatment history.

What is Machine Learning?

Machine learning, an application of artificial intelligence(AI),has some impressive capabilities. A machine learning algorithm can make software capable of unsupervised learning Without being explicitly programmed, the algorithm can seemingly grow "smarter," and become more accurate at predicting outcomes, through the input of historical data.

How machine learning is useful in disease prediction?

Disease Prediction using Machine Learning is the system that is used to predict the diseases from the symptoms which are givenby the patients oranyuser. The system processes the symptoms provided by the user as input and gives the output as the probability of the disease.

IT IS POSSIBLE TO SAVE MILLIONS OF LIVES THROUGH DISEASE PREDICTION USING MACHINE LEARNING
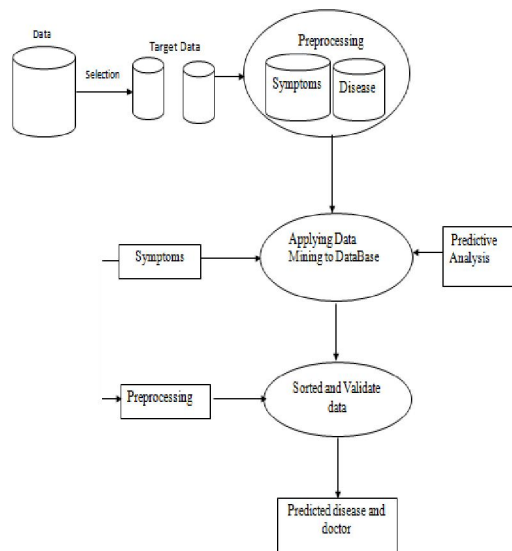
II. METHODOLOGY

Fig: Architecture of Propos System

Data is in the form of text documents, images, audio clips, software programs, or other types of data. The term data has been used to describe data in the petabyte range or larger. Data is in the form of unstructured manner.

Data selection is defined as a process where data related to the analysis is decided and retrieved data from the data collection. Data selection can be done using data mining technique as shown in the introduction.

Target Data is defined as the removal of noisy and irrelevant data from a collection. Cleaning in case of missing values. Cleaning noisy data, where noise is a random or variance error. After removal of unstructured data structured data is target data.

Data preprocessing is a data mining technique that involves transforming raw data into an understandable, structured format. In proposed system symptoms and disease can be added in a structured format.

In the healthcare system, a large number of a database is arranged in noisy format. In the proposed system, we can add symptoms using patient condition. This symptom stored in symptoms database.

III. OBJECTIVE

- The major objective is to analyze the symptoms given by the patients and predict the diseases.
- Based on the predicted disease proposed system recommend a doctor.
- Clustering and prediction algorithms used to predict disease and to suggest doctors

IV. SCOPE

The proposed system predicts the diseases are polio, dengue, lung disease, blood cancer.

This proposed system work can be used in medical sectors. The future scope of this research is to suggest nearest hospital

V. RESULT
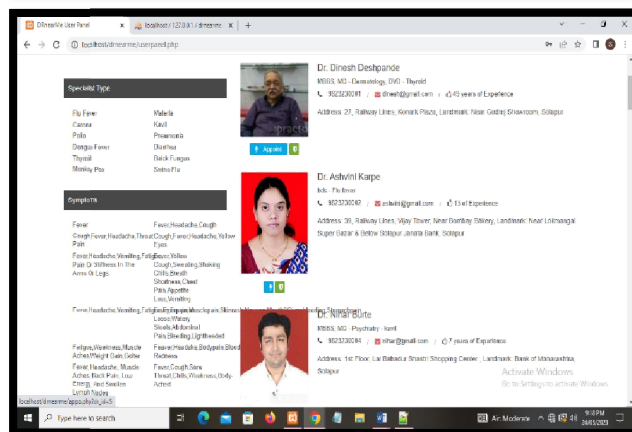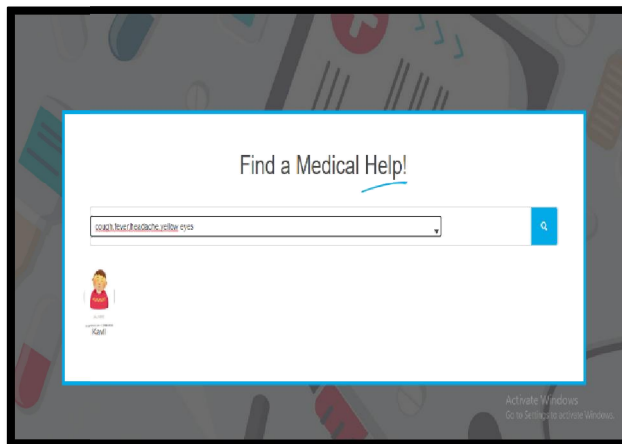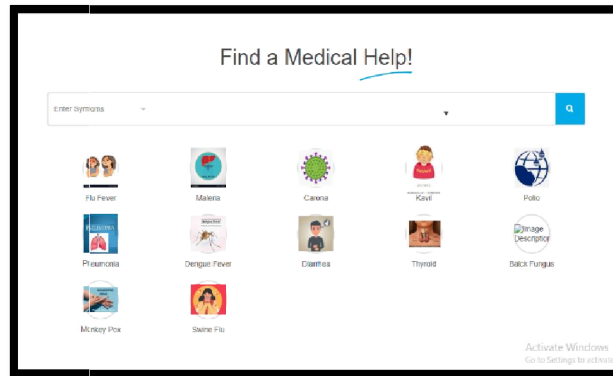
VI. ADVANTAGES

- Doctor can get more patient.
- Instant doctor help
- Quick and accurate medicine available

VII. CONCLUSION

We developed a Prediction Engine which enables the user to check whether he/she has diseases. The user interacts with the Prediction Engine by filling a form which holds theparameter set provided as an input to the trained models. The Prediction engine provides an optimal performance compared to other state of art approaches. The Prediction Engine makes use of three algorithms to predict the presence of a disease namely: Support Vector Machine (SVM), K-Nearest Neighbours (KNN) and Naïve Bayes. The reason to choose these three algorithms are:

- They are effective, if the training data is large.
- A single dataset can be provided as an input to all these 3 algorithms with minimal or no modification.
- A common scalar can be used to normalize the input provided to these 3 algorithms.
- Mail the detailed report of the prediction engine results along with the information of 5 nearest medical facilities details having location and contact information.
- Automate the process of training the model and extracting pickle files of the trained models which will be consumed by the API's to predict the disease

REFERENCES

[1] Prashanta Kumar Patra " Smart Health Care System using Data Mining "IEEE 2018, DOI 10.1109/ICIT.2018.00021

[2] Xiangxiang Zeng, Senior Member, Yinglai Lin, Yuying He, Linyuan L¨ u, Xiaoping Min_, and Alfonso Rodr´ıguez-Pat´ on " Deep collaborative filtering for prediction of disease genes " IEEE 2018

[3] Faezeh Movahedi1, Yiye Zhang2, Rema Padman3, James F. Antaki4, "Mining temporal patterns from sequential healthcare data " 2018 IEEE International Conference on Healthcare Informatics

[4] Khaled A. S. Abu Daqqa, Ashraf Y. A. Maghari, Wael F. M.AlSarraj "Prediction and Diagnosis of Leukemia Using Classification Algorithms" 2017 8th International Conference on Information Technology (ICIT).

[5] Dr.Shailendra Narayan Singh, Monika Gandhi " Predictions in Heart Disease Using Techniques of Data Mining " IEEE 2015, 978-1-4799-8433- 6/15/$31.00

[6] AnkitaDewan, Meghna Sharma "Prediction of Heart Disease Using a Hybrid Technique in Data Mining Classification" in 2015 IEEE

[7] Anu Sharma1, Dr. M.K Sharma2 &Dr. R.K Dwivedi3, " Literature Review and Challenges of Data Mining Techniques for Social Network Analysis " Advances in Computational Sciences and Technology ISSN 0973-6107 Volume 10, Number 5 (2017) pp. 1337-1354 © Research IndiaPublications http://www.ripublication.com

Copyright to IJARSCT
www.ijarsct.co.in

DOI: 10.48175/568

455

ISSN
2581-9429
IJARSCT