



Multiple Disease Prediction Using Different Machine Learning Algorithms Comparatively

Rudra A. Godse¹, Smita S. Gunjal², Karan A. Jagtap³, Neha S. Mahamuni⁴, Prof. Suchita Wankhade⁵

Student, Department of Computer Engineering, Trinity College of Engineering and Research, Pune, India^{1,2,3,4}

Assistant Professor, Department of Computer Engineering, Trinity College of Engineering and Research, Pune, India⁵

Abstract: Using Machine learning, our project proposes disease prediction system. For small problems, the users have to go personally to the hospital for check-up which is more time consuming. Also handling the telephonic calls for appointments is quite hectic. Such a problem can be solved by using disease prediction application by giving proper guidance regarding healthy living. Over the past decade, the use of the specific disease prediction tools along with the concerning health has been increased due to a variety of diseases and less doctor-patient ratio. Thus, in this system, we are concentrating on providing immediate and accurate disease prediction to the users about the symptoms they enter along with the severity of disease predicted. Best suitable algorithm and doctor consultation will be given in this project. For prediction of diseases, different machine learning algorithms are used to ensure quick and accurate predictions. In one channel, the symptoms entered will be crosschecked with the database. Further, it will be preserved in the database if the symptom is new which its primary work is and the other channel will provide severity of disease predicted. A web/android application is deployed for user for easy portability, configuring and being able to access remotely where doctors cannot reach easily. Normally users are not aware about all the treatment regarding the particular disease, this project also looks forward to providing medicine and drug consultation of disease predicted. Therefore, this arrangement helps in easier health management.

Keywords: Machine Learning, KNN algorithm, SVM, Decision Tree Algorithm, Naïve Bayes Algorithm, Django, Python, etc

I. INTRODUCTION

The Earth is passing through a purplish patch of technology, where there is increasing demand of intelligence and accuracy behind it. Today's people are more likely addicted to Internet but they are not concerned about their personal health. In this 21st Century humans are surrounded with technology as they are the constituent of our day to day life cycle. With this we are always focusing on the health for ourselves and our earned valuables respectively. People avoid to go in hospital for small problem which may become a major disease in future. Establishing question answer forums is becoming a simple way to answer those queries rather than browsing through the list of potentially relevant document from the web. Our basic idea is to develop a system which will predict and give the details of the disease predicted along with its severity which as symptoms are given as input by the user. The system will compare the symptoms with the datasets provided in the database. If the symptom matches the datasets then it should ask other relevant symptoms specifying the name of the symptom. If not, the symptom entered should be notified as wrong symptom. After this a prompt will come up asking whether you want to still save the symptom in the database. If you click on yes, it will be saved in the database, if not it will go to the recycle bin. The main feature will be the machine learning, in which we will be using algorithms such as Naïve Bayes Algorithm, K-Nearest Algorithm, Decision Tree Algorithm, Random Forest Algorithm and Support Vector Machine, which will predict accurate disease and Also, will find which algorithm gives a faster and efficient result by comparatively-comparing.

II. LITERATURE SURVEY

1. "Prediction of Cardiovascular Disease Using Machine Learning Algorithms" (2018). This paper contributes the correlative application and analysis of distinct machine learning algorithms in the R software which gives an immediate mechanism for the user to use the machine learning algorithms in R software for forecasting the cardiovascular diseases.
2. "A Proposed Model for Lifestyle Disease Predict Vectorion Using Support Machine" (2018). This study aims to understand support vector machine and use it to predict lifestyle diseases that an individual might be susceptible to.



3. “Multi Disease Prediction Using Data Mining Techniques” (2017).
In this study two different data mining classification techniques was used for the prediction of various diseases and their performance was compared in order to evaluate the best classifier. An important challenge in data mining and machine learning areas is to build precise and computationally efficient classifiers for Medical applications.
4. “Prediction of Heart Disease Using Machine Learning Algorithms” (2018).
In this paper, two supervised data mining algorithm was applied on the dataset to predict the possibilities of having heart disease of a patient, were analyzed with classification model namely Naïve Bayes Classifier and Decision tree classification. The Decision tree model has predicted the heart disease patient with an accuracy level of 91% and Naïve Bayes classifier has predicted heart disease patient with an accuracy level of 87%.
5. “Analysis of Heart Disease Prediction Using Datamining Techniques”(2017)
Heart disease is one of the leading causes of deaths worldwide and the early prediction of heart disease is very important. In this study prove that the proposed new algorithm achieves a highest accuracy compare with another algorithm.
6. “Review of Medical Disease Symptoms Prediction Using Data Mining Technique”(2017)
In this paper evaluate the performance of medical disease prediction based on data mining technique. The classifier classified the medical diagnosis of disease data such as cancer, liver problem, and heart disease and so on. SVM method better classified data in compression of conventional cluster ensemble technique.

III. EXISTING SYSTEM

In the existing system the data set is typically small, for patients and diseases with specific conditions. These systems are mostly designed for the more colossal diseases such as Heart Disease, Cancer etc. The pre-selected characteristics may sometimes not satisfy the changes in the disease and its influencing factors which could lead to inaccuracy in results. As we live in continuously evolving world, the symptoms of diseases also evolve over a course of time. Also most of the current systems make the users wait for long periods by making them answer lengthy questionnaires.

IV. PROPOSED SYSTEM

We are proposing such a system which will flaunt a simple and elegant User Interface and also be time efficient. In order to make it less time consuming we are aiming at a more specific questionnaire which will be followed by the system. Our aim with this system is to be the connecting bridge between doctors and patients. The main feature will be the machine learning, in which we will be using algorithms such as Naïve Bayes Algorithm, K-Nearest Algorithm, Decision Tree Algorithm, Random Forest Algorithm and Support Vector Machine, which will help us in getting accurate predictions and Also, will find which algorithm gives a faster and efficient result by comparatively-comparing. Another feature that our system will comprise of is Doctor’s Consultation. After delivering the results, our system will also suggest the user to get a doctors consultation on this report.

By using this feature, we will not only address the other class of users i.e. the Doctors but we will also gain their trust in this system as in that this system is not affecting their business.

V. ML SYSTEM ARCHITECTURE

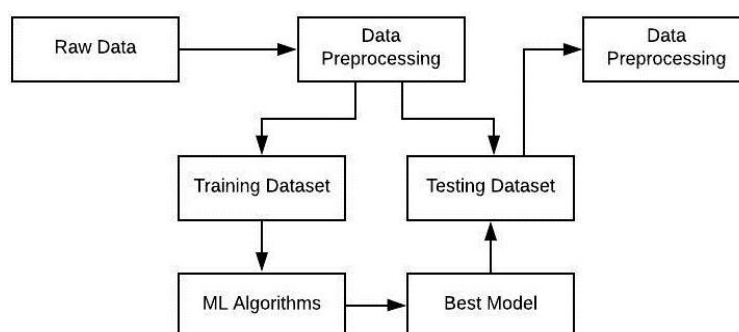


Fig.1 ML System Architecture

As shown in the above figure, the raw data from the original dataset is passed onto the first phase i.e. Data pre-processing. In Data pre-processing this raw data is then cleaned of all redundancies, missing values etc. The new clean data is fit for training different algorithmic models on it.



The process of training models is fundamental process in Machine learning Projects. There are two approaches to machine learning mainly Supervised Learning and Unsupervised Learning. Our model mostly applies the first approach initially. i.e. Supervised Learning.

Now in Supervised Learning, the system is trained on some examples i.e. Training set and then the model is asked to predict new values based on the test set.

The partitioning of dataset becomes crucial for getting good accuracy in models. The percentage mostly used while partitioning is 80/20 .i.e. 80% for training and 20% for testing purposes.

In our system we aim at first applying different algorithms on the training dataset and based on the model's Confidence and testing dataset accuracy, we select the best model algorithm and apply it on testing dataset to generate accurate results.

VI. CONCLUSION

This paper gives research of multiple researches done in this field. Our Proposed System aims at bridging gap between Doctors and Patients which will help both classes of users in achieving their goals. This system provides support for multiple disease prediction using different Machine Learning algorithms. The present approach of many systems focuses only on automating this process which lacks in building the user's trust in the system. By providing Doctor's recommendation in our system, we ensure user's trust side by side ensuring that the Doctor's will not feel that their Business is getting affected due to this System.

REFERENCES

- [1]. Jaymin Patel, Prof. Tejal Upadhyay, Dr. Samir Patel "Heart disease prediction using Machine learning and Data Mining Technique" Volume 7. Number 1 Sept 2015 March 2016.
- [2]. G. Parthiban, S.K. Srivasta "Applying Machine learning methods in Diagnosing Heart disease for Diabetic Patients" International Journal of Applied Information Systems (IJ AIS) – ISSN: 2249-0868 Foundation of Computer Science FCS, New York, USA Volume 3– No.7, August 2012.
- [3]. Thenmozhi.K and Deepika. P, Heart Disease Prediction using classification with different decision tree techniques. International Journal of Engineering Research & General Science, Vol 2(6), pp 6-11, Oct 2014.
- [4]. "Disease Prediction Using Machine Learning Over Big Data" Vinitha S, Sweetlin S, Vinusha H and Sajini S (2018).
- [5]. "Heart Disease Prediction using Logistic Regression Algorithm using Machine Learning" Reddy Prasad, Pidaparathi Anjali, S. Adil, N. Deepa (2019)
- [6]. "Heart Disease Prediction Using Data Mining Techniques" H. Benjamin Fredrick David and S. Antony Belcy (2018).
- [7]. "Multi Disease Prediction Using Data Mining Techniques" K. Gomathi, Dr. D. Shanmuga Priyaa (2017).