

Disease Prediction Using Machine Learning

Gaurav Shilimkar, Gaurav Shilimkar, Shivam Pisal

Department of Computer Engineering, Vishwakarma institute of technology, Pune, Maharashtra, India

ABSTRACT

Article Info

Volume 8, Issue 3

Page Number : 551-555

Publication Issue

May-June-2021

Article History

Accepted : 03 June 2021

Published : 08 June 2021

Big data has a significant part in a number of businesses, but it is largely essential to the rapidly growing healthcare industry. It plays an important role by offering a large set of data points, constructing a robust system which allows for better and more accurate results in disease detection. Originally, the forecasts are made on the information accessible, but the absence of imperfect information contributes to a decrease in the caliber of precision. Besides incomplete data different qualities of particular regional diseases, which change based on their areas of origin can weaken the prediction models further. In this paper we use data mining techniques such as association rule mining, classification, clustering and finally the Decision Tree Machine learning algorithm to analyze the different kinds of general body-based illnesses. We implemented and assessed the efficacy of the Decision Tree algorithm over real-life clinical information.

Keywords : Machine Learning, Precision, Information

I. INTRODUCTION

It is projected that, every 2 months, over 70% of the population in India has a tendency toward general body ailments like viral influenza, cough, and cold. Etc. Since a lot of people don't realize that the symptoms of these regular illnesses may be symptoms into something more detrimental, 25% of the population succumbs to death due to ignorance of the early signs. Hence, the identification of the disease in the initial stages is crucial for the prevention of any unwarranted casualties. The current medical system is mainly devoted to very specific, known diseases and is largely unequipped to identify and accurately predict diseases based on early signs.

The purpose of our system is to make predictions for the general and more commonly occurring disorder that when unchecked can become fatal diseases. The system applies data mining techniques, does pre-processing on the data and then implements the Machine Learning, decision tree algorithms. This system will predict the potential disorder based on the presented symptoms and precautionary measures required to avoid the aggravation of the condition, it will also aid the physicians analyze the patterns of current commonly occurring diseases. Within this project, the disease prediction system will execute data mining within its preliminary stages, the system will be trained and tested using decision trees.

This paper is split into five sections. The first section gives a brief introduction of concerning the system that we have chosen to implement. The second section is about data mining and the analysis of related other similar systems that have been previously executed. The third section details from the design of our system. The fourth section provides the results obtained. In the final section, the conclusion provides the overview and future scope of this system.

a long way to go before becoming an effective and integral part of the standard toolset that a typical software engineer uses day-to-day.

II. LITERATURE REVIEW

Literature reviews are a foundation for study in virtually every academic discipline. Hence, in this section we will present the current knowledge, substantive findings and existing systems have been successfully implemented and provided concrete evidence through their study.

III. METHODOLOGY

In this project we have chosen to work with the Machine learning algorithms and one of them was developed by Ross Quinlan. The algorithm is used to produce a decision tree from the specified dataset, works mostly on three things, firstly, the entropy of every feature, second, is the information gain and third, entropy of entire dataset, with these three, it chooses a root node. The requirement for the selection of a root node is that the feature with lower value of entropy or higher value of information gain becomes the root node. This carries on till the final element of the data that could offer some significant information isn't utilized.

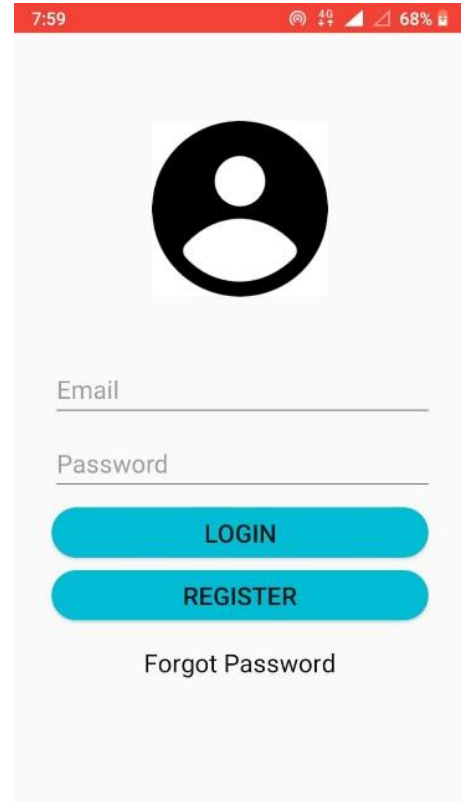
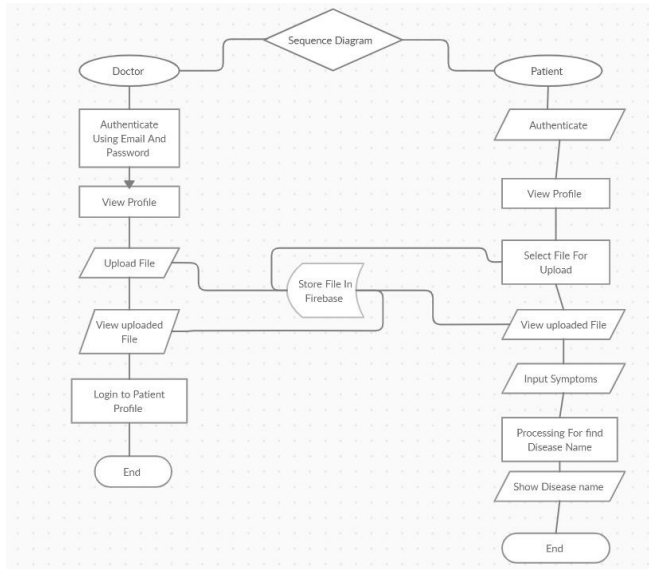
Identify applicable funding agency here. If none, delete this text box.

Following are the steps involved in the creation of a model using the Decision tree algorithm:

1. Calculation of the entropy of every attribute/feature using the dataset.
2. Splitting the dataset into subsets using the feature for which the resulting entropy (after splitting) is minimum (or, equivalently, information gain is maximum)
3. Making a decision tree root node using that feature.
4. Recursing on the subsets using remaining features until a full tree is constructed.

IV. System Design

The overall disease prediction system forecasts chance of existence of a disease within an individual on the grounds of their existing symptoms. It will then recommend precautionary measures needed to take care of the disease that is predicted. First, the system starts out by getting information fed to it from various sources i.e., patients, the information will then be pre-processed before additional procedure is performed, this is done in order to get clean data since the raw data is usually noisy, or faulty. This information will be processed using Data mining algorithms, and then the system will be trained in order to predict the disorder based on the input given by the user.



Step 1: Data Preparation

The system is implemented in two components, the admin section and the consumer part. The responsibility of the admin is training the machine for development of this disease prediction model. The user utilizes the services provided by the model after logging viz. inputting the symptoms to the design, which subsequently yields the predicted outcome and necessary precautionary steps.

V. IMPLEMENTATION

The system consists of two components: (i). The Admin (ii). The User

Admin:

Upon logging into the system as a verified admin, the admin performs the following tasks:

- Data Preparation.
- Data Transformation.
- Feature Extraction.
- Implementation of Machine learning algorithms.
- Model Creation
- Viewing Reports/documents

The data in this study was obtained from a local hospital based in Pune. The initial data consisted of 120 unique data points.

Step 2: Data Transformation

Within this step, the dataset is explored, and required data filtered and chosen. The data is then converted to the appropriate format that can be interpreted by the machine learning model.

Step 3: Feature Extraction

Feature extraction is the process of decreasing the size of data in order to only take informative non-redundant and relevant data, in order to ease subsequent learning and extrapolate it to the generalization step hence achieving better interpretation of results.

Step 4: Implementation of Machine Learning algorithms

For the purpose of this project, we have selected Machine Learning algorithms for training the disease prediction system. After a set of algorithms is applied, it creates a rule set based on the patterns that it identifies in the data that is fed to it. The system training is done based on this rule set and the model is generated.

Step 5: Model Creation

After the system training is complete, the disease prediction model will be ready for use.

Step 6: Viewing Reports/documents

Within this step we have added the feature for a doctor to access the documents of a patient for faster response

User:

On logging into the system as user, the user can perform the following functions on the system:

- Enter Symptoms
- Obtain Disease Predictions
- View Precautions
- Add documents/reports

Step 1: Enter Symptoms

After logging into the system, the user can select the symptoms that they need to check for, from those available in the drop-down menu.

Step 2: Obtain Disease prediction

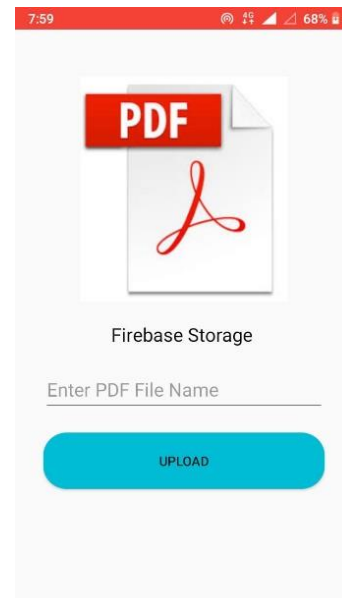
The system will then predict the disease that a person might have based on the symptoms selected by the user.

Step 3: View Precautions

Along with the prediction of the disease, the system displays the precautions to be taken in order to recover from the predicted disease.

Step 4: Add documents/reports

In this step the patient can add any of their reports/documents for the doctor to see, for faster response



VI. Results

The overall disease prediction system performs data mining and the algorithm that is the crux of the system are the Machine learning algorithms. Decision trees are fairly easy models to comprehend as the reason for each decision that is made is provided making it a logical model even for human understanding. Knowledge models below this paradigm can be directly transformed to some IF-THEN rules which are among the most common kinds of knowledge representation.

VII. Conclusion

The present system covers only the general illnesses or the commonly occurring diseases, the future scope for this project would be to include a broader range of diseases, especially those with a higher fatality rate, such a various kind of cancers, so that early detection could be achieved subsequently leading to the patient receiving medical attention much sooner. The fatality rate for many cancers reduces significantly if detected in the early stages and treatment is administered immediately.

Cite this article as :

Gaurav Shilimkar, Gaurav Shilimkar, Shivam Pisal , " Disease Prediction Using Machine Learning", International Journal of Scientific Research in Science and Technology(IJSRST), Print ISSN : 2395-6011, Online ISSN : 2395-602X, Volume 8, Issue 3, pp.551-555, May-June-2021. Available at doi : <https://doi.org/10.32628/IJSRST12183118> Journal URL : <https://ijsrst.com/IJSRST12183118>

VIII. REFERENCES

- [1]. Implementing WEKA for medical data classification and early disease prediction. "3rd IEEE International Conference on "Computational Intelligence and Communication Technology" (IEEE-CICT 2017)".
- [2]. Dr. B.Srinivasan, K.Pavya, "A study on data mining prediction techniques in healthcare sector", in International Research Journal of Engineering and Technology (IRJET), March-2016.
- [3]. Analysis of Data Mining Techniques for Heart Disease Prediction, "Marjia Sultana, Afrin Haider and Mohammad Shorif Uddin".
- [4]. Feixiang Huang, Shengyong Wang, and Chien-Chung Chan, "Predicting Disease By Using Data Mining Based on Healthcare Information System" , in IEEE 2012.