

Medical Diagnosis Using Machine Learning

Dr. V. S. Anita Sofia¹, Rahul Prasad.M²

¹ Associate Professor, Dept of MCA

²Dept of MCA

^{1,2} PSG College of Arts & Science, Coimbatore, India

Abstract- *Medical Diagnosis using Machine Learning is a system which predicts the disease based on the information or the symptoms he/she enter into the system and provides the accurate results based on that information. If the patient is not much serious and the user just wants to know the type of disease, he/she has been through. It is a system which provides the user the tips and tricks to maintain the health system of the user and it provides a way to find out the disease using this prediction. Now a day's health industry plays major role in curing the diseases of the patients so this is also some kind of help for the health industry to tell the user and also it is useful for the user in case he/she doesn't want to go to the hospital or any other clinics, so just by entering the symptoms and all other useful information the user can get to know the disease he/she is suffering from and the health industry can also get benefit from this system by just asking the symptoms from the user and entering in the system and in just few seconds they can tell the exact and up to some extent the accurate diseases. This Medical Diagnosis Using Machine Learning is completely done with the help of Machine Learning and Python Programming language with Tkinter Interface for it and also using the dataset that is available previously by the hospitals using that we will predict the disease.*

I. INTRODUCTION

Medical Diagnosis using Machine Learning is a system which predicts the disease based on the information provided by the user. It also predicts the disease of the patient or the user based on the information or the symptoms he/she enter into the system and provides the accurate results based on that information. If the patient is not much serious and the user just wants to know the type of disease, he/she has been through. It is a system which provides the user the tips and tricks to maintain the health system of the user and it provides a way to find out the disease using this prediction. Now a day's health industry plays major role in curing the diseases of the patients so this is also some kind of help for the health industry to tell the user and also it is useful for the user in case he/she doesn't want to go to the hospital or any other clinics, so just by entering the symptoms and all other useful information the user can get to know the disease he/she is suffering from and the health industry can also get benefit

from this system by just asking the symptoms from the user and entering in the system and in just few seconds they can tell the exact and up to some extent the accurate diseases. This DPUML is previously done by many other organizations but our intention is to make it different and beneficial for the users who are using this system. This Medical Diagnosis Using Machine Learning is completely done with the help of Machine Learning and Python Programming language with Tkinter Interface for it and also using the dataset that is available previously by the hospitals using that we will predict the disease.

II. ALGORITHMS

DECISION TREE ALGORITHM: Decision tree induction is the learning of decision trees from class-labelled training tuples. A decision tree is a flowchart-like tree structure, Decision tree induction is a non-parametric approach for building classification models. Finding an optimal decision tree is an NP-complete problem. Techniques developed for constructing decision trees are computationally inexpensive, making it possible to construct models even when the training set size is very large. Decision trees, especially smaller-sized trees, are relatively easy to interpret. Decision tree provide an expressive representation for learning discrete- valued functions. Decision tree algorithms are quite robust to the presence of noise, especially when methods for avoiding overfitting.

RANDOM FOREST ALGORITHM: It is an ensemble classifier using many decision trees models; it can be used for regression as well as classification. Accuracy and variable importance information can be provided with the results. A random forest is the classifier consisting of a collection of tree structured classifiers k , where the k is independently; identically distributed random trees and each random tree consist of the unit of vote for classification of input. Random forest uses the index for the classification and determining the final class In each tree. The final class of each tree is aggregated and voted by the weighted values to construct the final classifier.

NAÏVE BAYES ALGORITHM: It is used to predict the categorical class labels. It classifies the class data based on the

training set and the values in a classifying attribute and uses it in classifying new data. It is a two-step process Model Construction and Model Usage. This Bayes theorem is named after Thomas Bayes and it is statistical method for classification and supervised learning method. It can solve both categorical and continuous values attributes. Bayes theorem finds the probability of an event occurring given the probability of another event that has already occurred. Bayes theorem is stated mathematically as the following equation.

III. SYSTEM ANALYSIS

Now a day's health industry plays major role in curing the diseases of the patients so this is also some kind of help for the health industry to tell the user and also it is useful for the user in case he/she doesn't want to go to the hospital or any other clinics, so just by entering the symptoms and all other useful information the user can get to know the disease he/she is suffering from and the health industry can also get benefit from this system by just asking the symptoms from the user and entering in the system and in just few seconds they can tell the exact and up to some extent the accurate diseases.

Existing System: Prediction using traditional methods and models involves various risk factors and it consists of various measures of algorithms such as datasets, programs and much more to add on. High-risk and Low-risk patient classification is done on the basis of the tests that are done in group. But these models are only valuable in clinical situations and not in big industry sector. So, to include the disease predictions in various health related industries, we have used the concepts of machine learning and supervised learning methods to build the predictions system.

Drawbacks In Existing System: Now a day's in Health Industry there are various problems related to machines or devices which will give wrong or unaccepted results, so to avoid those results and get the correct and desired results. The health industry in information yet and knowledge poor and this industry is very vast industry which has lot of work to be done.

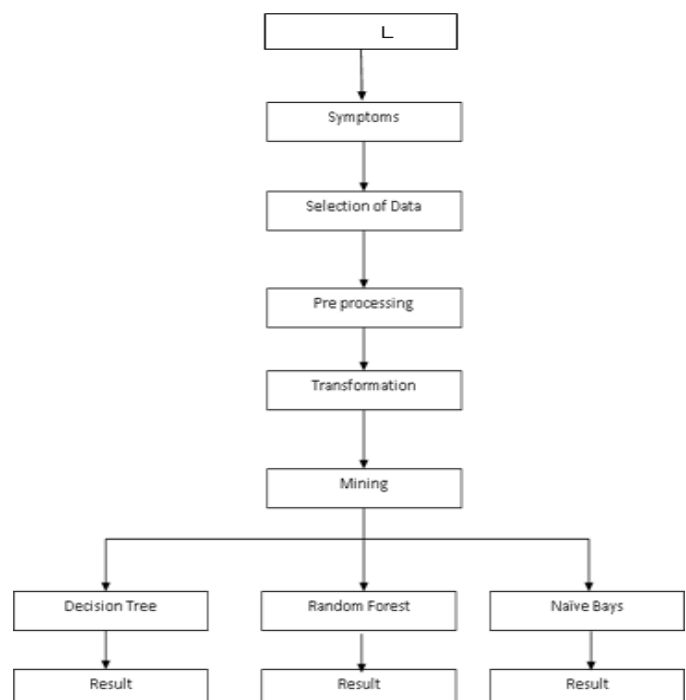
Proposed System: The proposed system of medical diagnosis using machine learning is that we have used many techniques and algorithms and all other various tools to build a system which predicts the disease of the patient using the symptoms and by taking those symptoms we are comparing with the system's dataset that is previously available. By taking those datasets and comparing with the patient's disease we will predict the accurate percentage disease of the patient. The dataset and symptoms go to the prediction model of the system where the data is pre-processed for the future

references and then the feature selection is done by the user where he will enter the various symptoms. Then the classification of those data is done with the help of various algorithms and techniques such as Decision Tree, Naïve Bayes, Random Forest and etc. Then the data goes in the recommendation model, there it shows the risk analysis that is involved in the system and it also provides the probability estimation of the system such that it shows the various probability like how the system behaves when there are n number of predictions are done and it also does the recommendations for the patients from their final result and also from their symptoms like it can show what to use and what not to use from the given datasets and the final results.

Objectives of Proposed System: This Project will predict the diseases of the patients based on the symptoms and other general information using the datasets. This is done based on the previous datasets of the hospitals so after comparing it can provide up to 80% of accurate results, and the project is still developing further to get the 100% accurate results. With the help of Disease prediction, it can predict the disease of the patient and can solve various problems and prevents from various aspects. It provides security for the system so that no one can break into that and no one can make any changes in the system.

IV. SYSTEM DESIGN

System flow diagram are a way of displaying how data flows in a system and how to decisions are made to control event



INPUT DESIGN: Input design is the process of converting user-originated inputs to a computer understandable format. Input design is one of the most expensive phases of the operation of computerized system and is often the major problem of a system. A large number of problems with a system can usually be tracked back to fault input design and method. Every moment of input design should be analyzed and designed with utmost care. The design of the input should be made the input as the over to the numerous networks in the reliable area that should be passed as the installation in the remote network. It has the following constraints in the input database. All the files from the disk should be acquired by data. It is suitable to more available data clearance and made available. The menu of design should be understandable and it is in the right format. The system takes input from the users, processes it and produces an output. Input design is link that ties the information system into the world of its users. The system should be user-friendly to gain appropriate information to the user. The project gives the low time consumption to make the sensitive application made simple. When applying the project it provides the low man-power attrition with the reasonable output. The amount of fund that the company can spend into the research and development of the system is limited. The expenditures must be justified. Thus the developed system as well within the budget and this was achieved because most of the technologies used are freely available. Only the customized products had to be purchased.

OUTPUT DESIGN: Output design generally refers to the results and information that are generated by the system for many end-users; it should be understandable with the enhanced format. The Output of the software is used to make the remote installation of the new software in the system and, it is awake the immediate alert to the system that should be enhanced it as the input to the system. Output is the main reason for developing the system and the basis on which they evaluate the usefulness of the application. Computer output is the most important direct source of information to the user output design deals with form design efficient output design should improve the interfacing with user. The term output applies to any information produced by an information system in terms of displayed. When analyst design system output, they Identify the specific output that is needed to meet the requirements of end user. Previewing the output reports by the user is extremely important because the user is the ultimate judge of the quality of the output and, in turn, the success of the system When designing output, system analysis accomplishes more things like, to determine what applications, websites or documents whether blocked or allowed. Allowing should be in various options. The output is designed in such a way that it is attractive, convenient and informative.

V. TESTING & IMPLEMENTATION

After the source code has been completed, documented as related data structures. Completion of the project has to undergo testing and validation where there is subtitle and definite attempt to get errors. The project developer treats lightly, designing and execution of the project test that will demonstrates that the program works rather than uncovering errors, unfortunately errors will be present and if the project developer doesn't found errors, the user will find out. The project developer is always responsible for testing the individual units

i.e. modules of the program. In many cases, developer should conduct the integration testing i.e. the testing step that leads to the construction of the complete program structure.

Unit Testing: Unit testing focuses verification efforts on the smallest unit of software design of the module. This is also known as "module testing". This testing is carried out during programming stage itself. In this testing step, each module is found to be working satisfactorily as regards to the expected output of the modules.

Integration Testing: Data can be lost across an interface, one module can have adverse effect on another sub function when combined it may not produce the desired major functions. Integration testing is a systematic testing for constructing test to uncover errors associated within an interface. The objectives taken from unit tested modules and a program structure is built for integrated testing. All the modules are combined and the test is made. A correction made in this testing is difficult because the vast expenses of the entire program complicated the isolation of causes. In this integration testing step, all the errors are corrected for next testing process.

Validation Testing: After the completion of the integrated testing, software is completely assembled as a package; interfacing error has been uncovered and corrected and a final series of software test validation begins. Validation testing can be defined in many ways but a simple definition is that validation succeeds when the software function in a manner that can be reasonably expected by the customer.

VI. CONCLUSION AND FUTUR SCOPE

So, Finally conclude by saying that, this project Medical Diagnosis using Machine Learning is very much useful in everyone's day to day life and it is mainly more important for the healthcare sector, because they are the one that daily uses these systems to predict the diseases of the

patients based on their general information and there symptoms that they are been through. If health industry adopts this project then the work of the doctors can be reduced and they can easily predict the disease of the patient. The Disease prediction is to provide prediction for the various and generally occurring diseases that when unchecked and sometimes ignored can turns into fatal disease and cause lot of problem to the patient and as well as their family members. To address these issues, we can create another classification that operate on the timeline itself, selecting tree splits based on intervals in the past. By modifying standard variable importance techniques for random forests, we are also able to present a system data of how the importance of a given variable varies over time, improving interpretability of the model. However, they don't fully take into account the temporal nature of system data. Also, while individual trees are relatively human interpretable, the weighted combination of dozens or hundreds of trees is not. The problem is exacerbated by the fact that the trees in random forest are typically not pruned, often resulting in trees with larger depth than those learned by decision tree algorithms like ID3 and C4.5

- Facility for modifying user detail.
- More interactive user interface.
- Facilities for Backup creation.
- Can be done as Web page.
- Can be done as Mobile Application.
- More Details and Latest Diseases.

REFERENCES

- [1] Disease Prediction and Doctor Recommendation System by www.irjet.net
- [2] Disease Prediction Based on Prior Knowledge by www.hcup-us.ahrq.gov/nisoverview.jsp
- [3] GDPS - General Disease Prediction System by www.irjet.net
- [4] Medical DiagnosisUsing Machine Learning by International Research Journal of Engineering and Technology (IRJET).
- [5] Kaveeshwar, S.A., and Cornwall, J., 2014, "The current state of disease mellitus in India". AMJ, 7(1), pp. 45-48.