**SMART HEALTH DISEASE PREDICTION WITH SYMPTOMS**

**ABSTRACT:** Decision tree is a powerful technique to classify the given data set and find the appropriate data that need to be discovered. The data set is obtained from Kaggle.com related to diseases and its corresponding symptoms are classified using decision tree, according to the decision tree algorithm. The model is trained with specific test data and a simple graphical user interface to collect the symptoms from the user and approximately predict the disease.

Key words: decision tree, interface, classification

1. **INTRODUCTION**

A generative technique for creating predictive modelling using specific cases is machine learning. It's a division of AI that advances the theory, that machines can see patterns in data and learn, make judgments using only the smallest amount of human intervention. A programming language is machine learning. algorithm that makes use of sample data or earlier gathered information to improve outcomes with high accuracy. The machine is in two stages. Learning process, planning and investigation. The logs of the user's or patient's symptoms used to foretell disease Learning Machines Technology provides a robust application environment in the addressing health illness prediction in the medical sector a user's or patient's experience-related concern.

In recent days, implementing machine learning in medical field is an active area for research, so that user can have a suggestion before consulting doctor, in accordance to that as a step this model will serve the purpose of predicting diseases of patients from symptoms that they are suffering. There are many methods to classify data like k-nearest neighbor which is based on selecting k nearest neighbor and classifying new data, Bayes’ theorem which is based on conditional probability, decision tree where particular set of data is selected as a based and all other data is divided based on that.

The method we choose for classification depends on dataset i.e., if data set is bigger or smaller. If the data set is considerably large, we use neural networks to train model

1. **MOTIVATION**

Medical field is an ever-green place for engineers to do research and help the medical requirements and pre requirements for complicated surgery. So as an initiative we had an idea to do a machine learning algorithm that can predict diseases using the symptoms from the user so that use can have an idea about disease and it can serve as a second opinion for user.

1. **Project Analysis**

**2.1. Objective**:

There is some sort of resources available to predict smart health. However, chronic diseases have been studied in particular and a level of risk has been identified. However, these methods are not widely used for disease prediction in general disease. Smart health prediction helps in the diagnosis of multiple diseases by analysing patient symptoms using a perfect fitting Machine Learning Algorithm technique.

**2.2. Existing Methods**:

The framework predicts chronic diseases for a specific area and population. Disease Prediction is for specific diseases only. In this method, Decision tree Algorithm is used to predict disease risk. The method uses Machine Learning algorithms such as K-nearest neighbours and Decision Tree. The machine has an accuracy value of 94.8 percent for some diseases. We are testing updated prediction models using real-world hospital data from certain specific regions/areas.

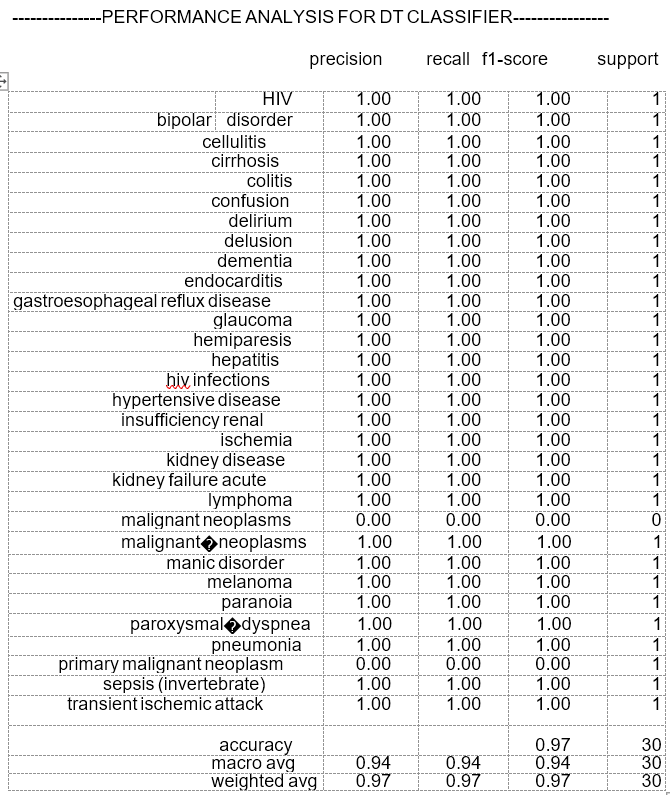


Fig 1: analysis for Decision tree



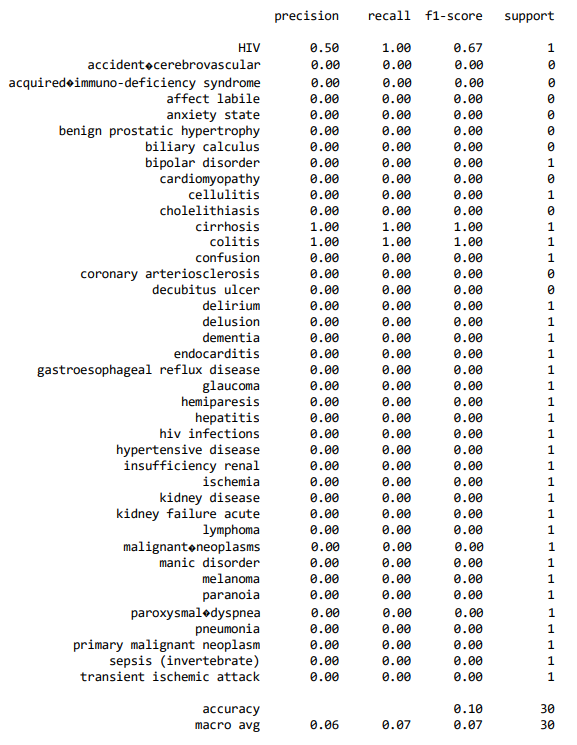
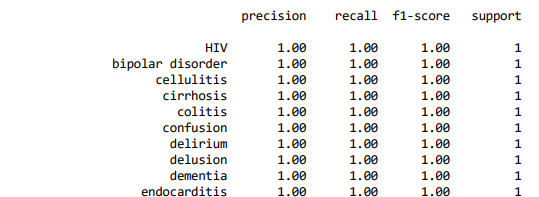


Fig 2: analysis of KNN classifier





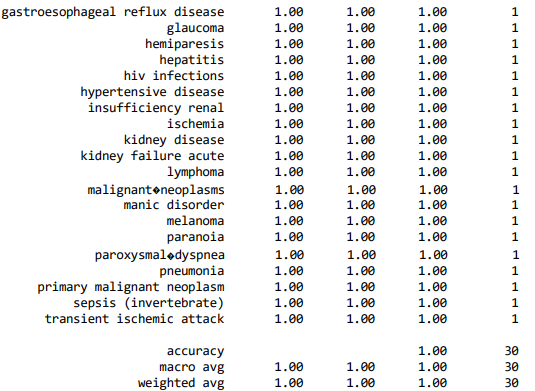
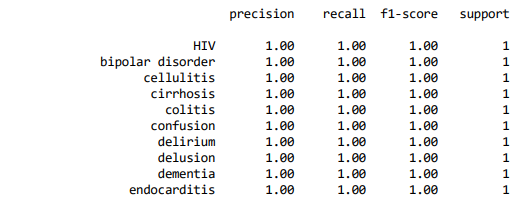


Fig 3: analysis of SVM classifier





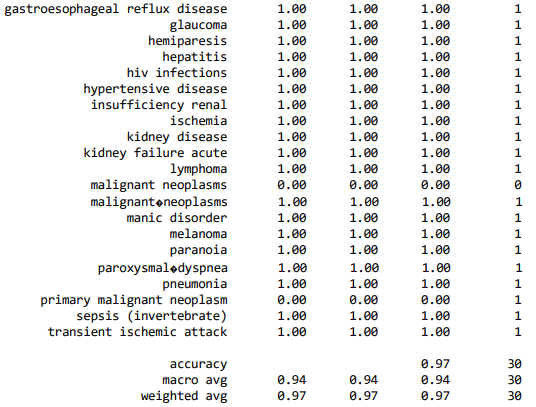


Fig 4: Analysis of Bayer’s classifier

Results of KNN classifier, SVM classifier, Decision tree classifier, Bayer’s classifier are shown above and among these Decision tree classifiers of selected and used for designing the model

**2.3 proposed method:**

Considering the above results, the proposed method of classification for this project is decision tree.

If someone is actually diagnosed with some sort of disease, they need to see a doctor/physician which is both time consuming and expensive too. It can also be difficult for the user to reach of doctors and hospitals so, the disease cannot be detected. Because, if the above procedure can be done with electronic software application that saves time and resources, it could be better for the patient to do the process runs smoothly. Smart health prediction is a web-based programme that predicts a user's illness based on their symptoms that the user/patient can feel. Data sets for the Smart Health Prediction Framework have been compiled from various health-related websites.

The consumer will be able to determine how likely a illness based on the symptoms displayed in the web-based programme. This project's goal is to

build a website that can forecast sickness

occurrences depending on several symptoms. Clients can select from a variety of symptoms to identify disease estimations using probability and conditions.

1. **METHODOLOGY**

3.1 Data set is downloaded which has disease names and its symptoms associated with it. When designing the algorithm, we assumed that the client would have a clear understanding of the signs he was seeing. The constructed prediction takes into account 95 manifestations, and the customer might accept the signs of his preparation as input throughout that time.

**Input dataset**

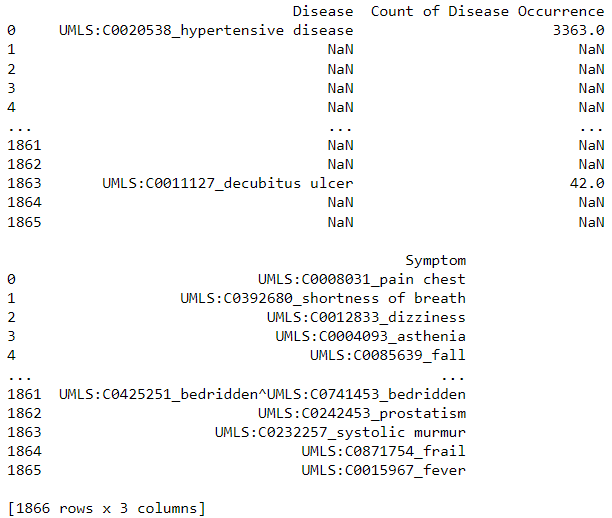


Fig 5: Input dataset

3.2 Pre-processing methods are applied to make nan data as usable data. Data pre-processing: The methods used in the mining of data that transform the raw data or re-encrypt it to create a structure so that it can be successfully decoded using computation are referred to as information pre-processing. The following list includes the information pre-processing techniques used in the work that was just introduced:

1. **Data purification:** This involves taking certain actions, such as adding back value that has been lost, to eliminate inconsistencies in the information.
2. **Data reduction**: When dealing with a large information base, the investigation becomes challenging. So, we exclude those independent variables (symptoms) that might not have an impact on the objective variables (diseases). Which of the approximately 95 of 132 adverse effects that are clearly associated with the illnesses will be picked for the ongoing task
3. **Models**: The entire system is built to predict diseases using three algorithms, namely the Decision Tree model and the Random Forest classifier model. This allows the predictive analysis study to be proposed at the end of the study by examining the speed, efficiency, and performance of the different algorithms for the input dataset.

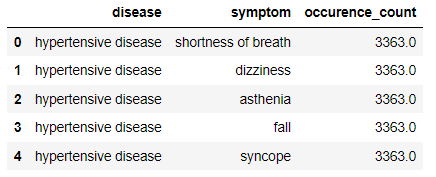


Fig 6: pre-processed input data set

3.3. Split data into training and test data for training model.

3.4. Train the model to decision tree algorithm

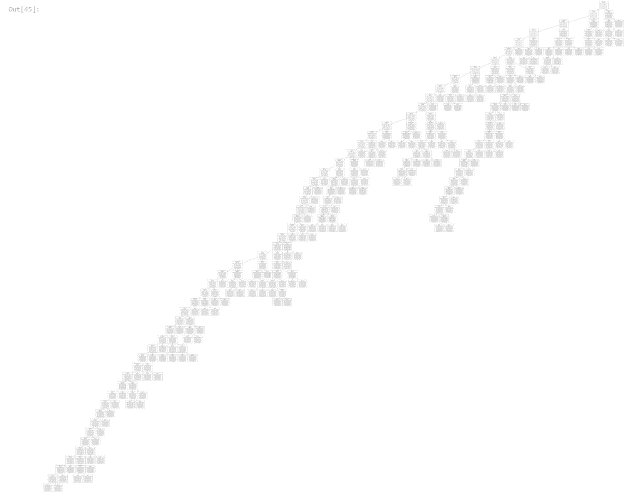


Fig 7: Decision Tree Classification

3.5 Use test data to check the model

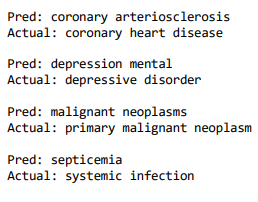


Fig 8 : testing model with test data

Input regarding an object is mapped to the item's output using the decision tree learning technique. Classification trees are tree models that have finite output classes. In these tree-like structures, the leaves represent class labels, while the branches depict the relationships between system class names and attribute values. Regression trees are decision trees with continuous output classes. A decision tree can be a decision-making input in data mining.

Finally, from the recorded advancement of (ML) Machine Learning technique and the approaches in clinical area, it very well may be shown that frameworks and systems have been arisen that has empowered refined information investigation by basic and direct utilization of Machine Learning (ML) models. This paper brings an extensive near investigation of three models’ execution of a clinical record with each of the obtaining accuracy score up to 98 %. Finally, the paper is investigated with disarray lattice & precision value. Man-made consciousness will be assumed significantly more significant part in information investigation later on because of the accessibility of gigantic information created and put away by the cutting-edge innovation.

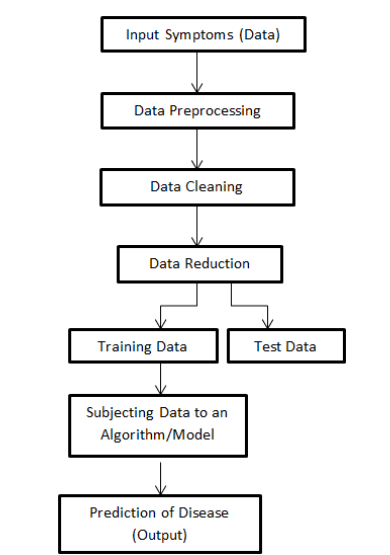


Fig 9: Overall methodology

**INTERFACE DESIGN:**

To get input from the user a user interface has been designed to get input symptoms from the user and model will print the output in the same interface window

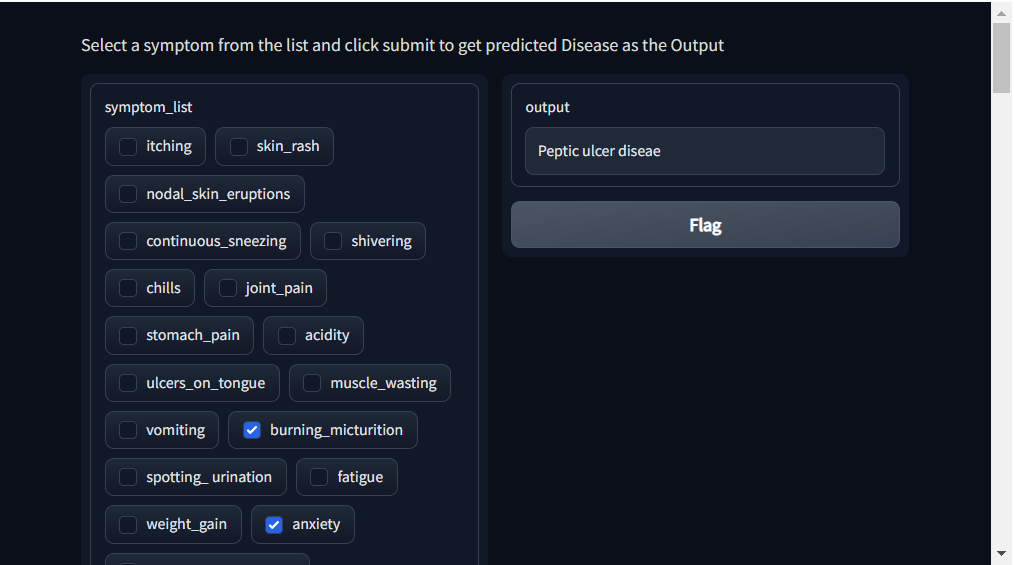
****

Fig 10: interface design

Interface design has been designed using ‘gradio’ library in python, which has inbuilt functions to make the interface interactable and this helps the user to select the symptoms that he is experiencing and result can also be printed.

1. **RESULTS**

When a framework is created with the preparation set and standard datasets are formed using validated calculations, whenever client indications are provided as an input to the algorithm, and the side effects are composed in agreement with the standard dataset created, this results in the creation of plans and the prediction of the highly probable infection.

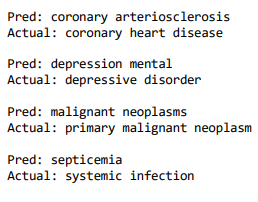


Fig 11 : testing model with test data

1. **FUTURE INPLEMENTATION**

The goal of this future implementation is to produce a web application forum for predicting disease manifestations on the basis of different symptoms and conditions. The user will pick different symptoms and find the diseases with their probabilistic data from the collected set of datasets.

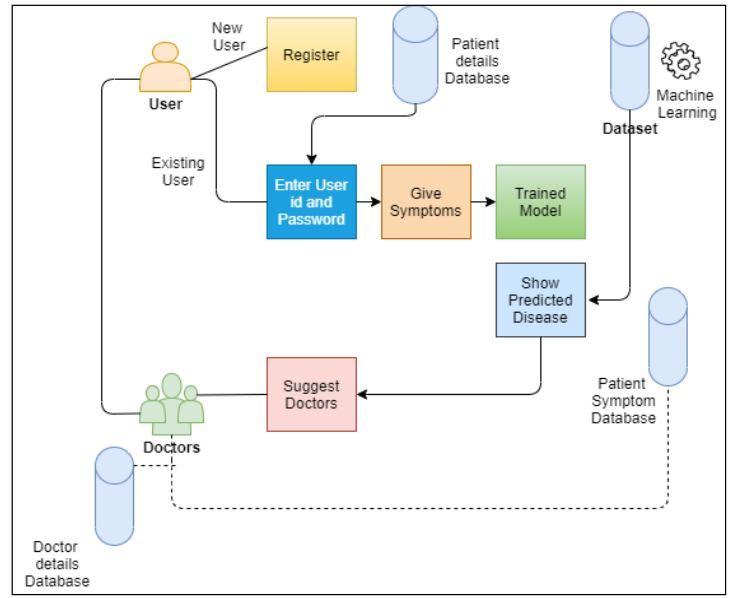
;

Fig 12: Future Implementation Idea

1. **CONCLUSION:**

Conclusion Required clinical symptom related information can be obtained from historical knowledge in the suggested methodology by planning datasets using the Naïve Bayes algorithm. Smart health can only be achieved if the system responds in this way. These datasets will be compared with the incoming queries and an Association Rule Mining Report will be generated. Given that this new solution will be based on real historical data, it would provide accurate and prompt results that would allow patients to get an urgent diagnosis. WebApplication such as sending a doctor remotely for a chat session are often provided so that patients can speak directly with physicians. As a result, in the true sense, this web system will be predictable and also produce high accuracy with fairness.