THINK INDIA JOURNAL

ISSN:0971-1260
Vol-22-Issue-10-November-2019

Diagonising a disease and connecting near by doctors

Lakshmi S V S S [1],M Sunanda[2], KP Akash[3],L Sumanth[4]

[1] Assistant Professor at Department of CSE, Anil Neerukonda Institute Of Technology And Sciences(A), Visakhapatnam-531162,India

[2,3,4] Final year students of Department of CSE, Anil Neerukonda Institute Of Technology And Sciences(A), Visakhapatnam-531162,India

Abstract

This paper represents the techniques related to data mining in medical field. The main objective of developing this project is to provide a proper medical guidance to the patient for their health issue with accurate results . This system reduces the tedious task of an applicant who apply manually by waiting for hours in the queue. Here, the system concentrates on the symptoms of patient's disease and based on the symptoms, the data is classified from the dataset and finally the disease name is predicted . After this, the patient can view various doctors and choose the best experienced doctor .This can be achieved by using data mining techniques which uses already existing data in different databases that transforms efficient results . Data mining is a process of extracting data from large amount of unused data into a meaningful information .Healthcare institutions mainly use data mining applications which has the main aspect of predicting future requests ,desires ,needs and conditions of the applicant to make optimal decisions about their treatments. Therefore, we conclude that data mining will achieve its full potential in the discovery of knowledge in medical field.

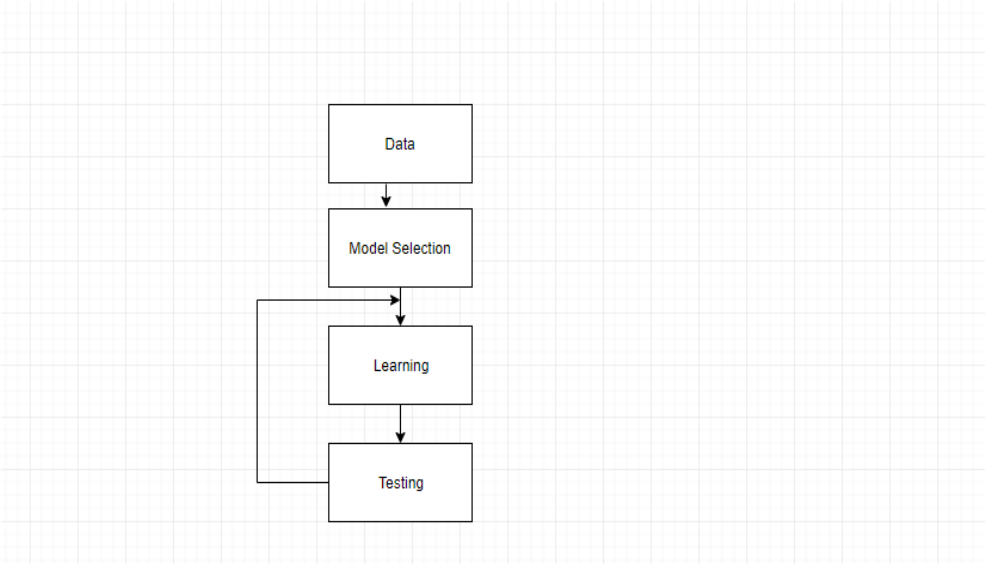
Key words : Health care ,Symptoms ,Disease ,Prediction ,Doctor , Applicant,Decision tree,Random Forest,Naive bayes .

Introduction

Human beings are getting affected by various diseases has become a common criteria now a days. They are affected by different diseases and neglecting the cause because of enlarged time taken to know the type of disease in their busy life and this is becoming a critical situation when it goes to a chronic stage. So , inorder to overcome the above mentioned problems we are developing this proposed system to provide an ease of access to the victims of the disease through online by saving their time. The two domains which we are using for this system is Data mining and machine learning.Machine Learning is defined as ability to learn.It comes under one of the best well-posed learning problems.A computer program is said to learn from Experience 'E' with respect to some class of Tasks 'T' and performance measure 'P'.It its performance at tasks in T is measured by P improves with Experience E.

ISSN:0971-1260
Vol-22-Issue-10-November-2019

Designing a Learning System

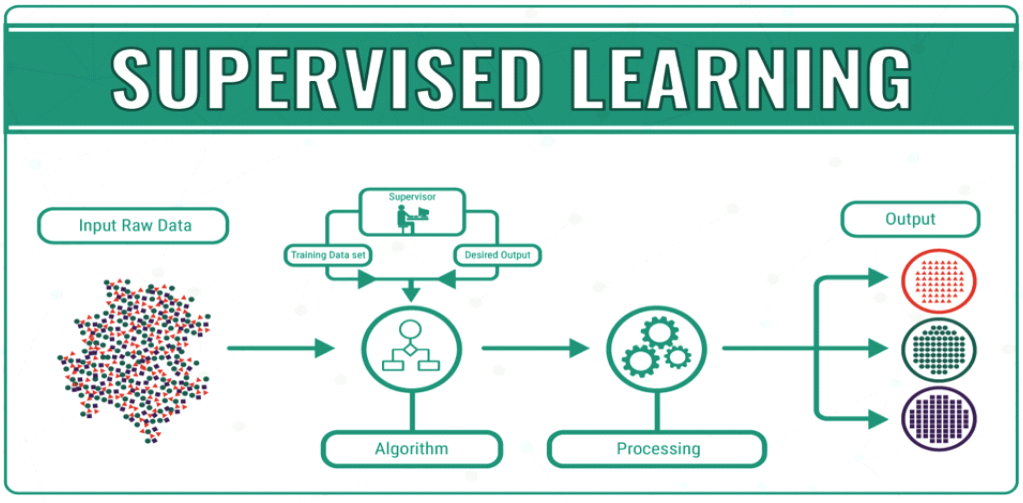


Machine Learning consists of two types of Learning.They are:

- Supervised Learning
- Unsupervised learning

Supervised Learning

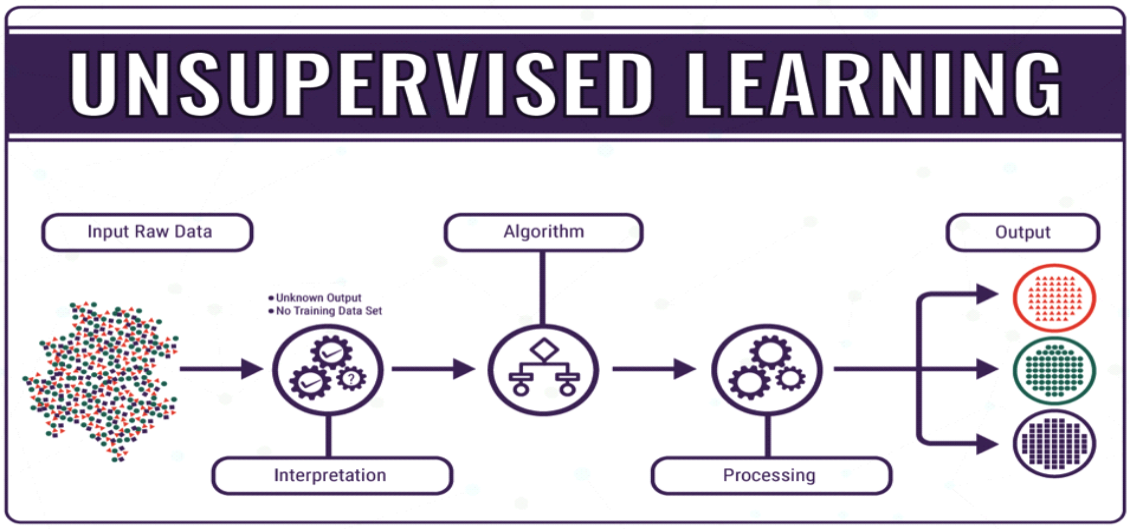
Learning something with the help of a supervisor or under some guidance comes under Supervised Learning.This learning process is completely dependent.Here,the input vector is presented to the network which will produce an output vector.The main aim is , the output vector and the target vector i.e, the desired output must be equal or else the adjustment of weights will continue till the outputs match with each other.



ISSN:0971-1260
Vol-22-Issue-10-November-2019

Unsupervised Learning

Learning is performed without the help of a guide or a teacher.Here,the network receives input patterns and organise it into a form of clusters.The network itself has the ability to learn by itsellf,discover patterns,features etc from input data .It is also called as Self-Organised Learning.



Literature Survey

In this system we are using naïve bayes theorem for classification of data and disease prediction.Naive Bayes classifiers are defined as a collection of classification algorithms based on **Bayes' Theorem**. The main priniciple is each instance is independent of each other .Naïve Bayes Algorithm is an example of supervised Learning i.e,it is completely trained by examples in the dataset.The most probable classification of new instance is obtained by combining the predictions of hypothesis.There are two events A and B where P (B) is true.P (A) is the **priori** of A (the prior probability, i.e. Probability of event before evidence is seen). The evidence is an attribute value of an unknown instance(here, it is event B).P (A|B) is a posteriori probability of B, i.e. probability of event after evidence is seen.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

1)**Naïve Bayes Algorithm**

Naïve Bayes Algorithm is a classification algorithm based on Bayes theorem's use in predictive modeling and this algorithm uses Bayesian techniques.This Algorithm is less computationally

ISSN:0971-1260
Vol-22-Issue-10-November-2019

intense then other and therefore is useful for quickly generating mining models to discover relationships between input and predictable columns.

Requirements for naïve bayes models:

Single key column: Each model must contain unique column which uniquely identifies each record.

Input columns: In Naive Bayes model, all columns must be discrete. It is also important to ensure that the input attributes are independent of each other.

At least one predictable column: The values of the predictable column can be treated as inputs.

2)Decision Tree Algorithm

It comes under one of the type of supervised learning algorithm which is especially used in classification problems.It works on both continuous as well as categorical input and output variables.In this,it consists of a single root node with homogeneous set of different nodes.

Decision tree is classified into 2 types.They are:

1)Categorical Variable Decision Tree:If a tree consists of a categorical target value then it is termed as Categorical variable decision tree.

2)Continuous Variable Decision Tree:If a tree consists of a continuous target variable then it is termed as Continuous Variable Decision tree.

Working of a Decision Tree:

A decision tree is a graphical representation of all possible solutions to a decision based on certain conditions.Each path of the tree is from root to leaf.

Decision tree learning is generally best suited to the problems with the following characteristics:

1)Instances are represented by attribute pair values.Attributes takes on a small number of disjoint possible values.

2)The target function is has discrete output values.Decision tree assigns a Boolean classification(Yes or No) to each example.

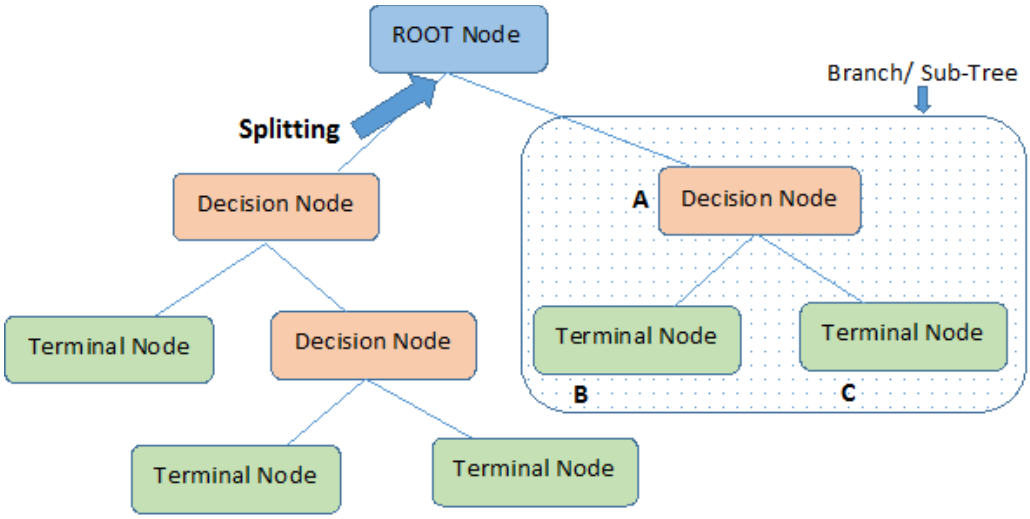
3)Disjunctive descriptions may be required.

4)Decision tree naturally represent disjunctive expressions.The training data may contain errors.

5)Decision tree learning methods are robust to errors.

6)The training data may contain missing attribute values.

ISSN:0971-1260
Vol-22-Issue-10-November-2019



Note:- A is parent node of B and C.

3)Random forest Algorithm:

Random forest algorithm is the term for an ensemble classifier which consists of many decision trees and therefore outputs the class that is the mode of the classes output by individual trees. Random forests are set of trees, all slightly different. It randomizes the algorithm, not the training data and generally improves the decisions of decision trees. Every decision tree will be constructed by using a Random subset of the training data from the dataset. To improve the predictive accuracy and control over-fitting, random forest acts as an estimator that fits various samples of dataset. Random Forests are especially used for predicting continuous variables .These are also used for estimating the probability that a particular outcome occurs, Outcomes can be either "yes/no" events i.e,True or false. It could have many possible outcomes but typically multi-class problems have 8 or fewer outcomes.

Imagine drawing a random sample from your main data base and building a decision tree on this random sample, This "sample" typically would use half of the available data although it could be a different fraction of the master data base. By repeating this process,we can draw a second different random sample and grow a second decision tree. The second decision tree makes different predictions which are typically different (at least a little) than those of the first decision tree.

Every tree generates its own specific predictions and acts independent for each record in the data base. By combining all of these separate predictions we can use the concept of voting or averaging. For predicting we would average the predictions made by our trees. By yes vs no percentage we can know the prediction.

Drawbacks of other algorithms:

One more algorithm is also there for implementing this framework namely Hill climbing algorithm but it has some drawbacks when compared to Naïve bayes algorithm and Random forest algorithm.The following are the disadvantages of hill climbing algorithm.

ISSN:0971-1260
Vol-22-Issue-10-November-2019

- It got failed to find a better solution.
- Either algorithm may terminate not by finding a goal state but by getting to a state from which no better state can be generated.
- Not an efficient method- It gets stuck at all peaks.

Methodology:

Features Available in the System

- Patient Registration & Login.
- View Details.
- Diseases Prediction.
- Admin Login
- Notification.
- Feedback
- Automatic Location Tracker

In this system there are two modules.They are:

Admin Module:

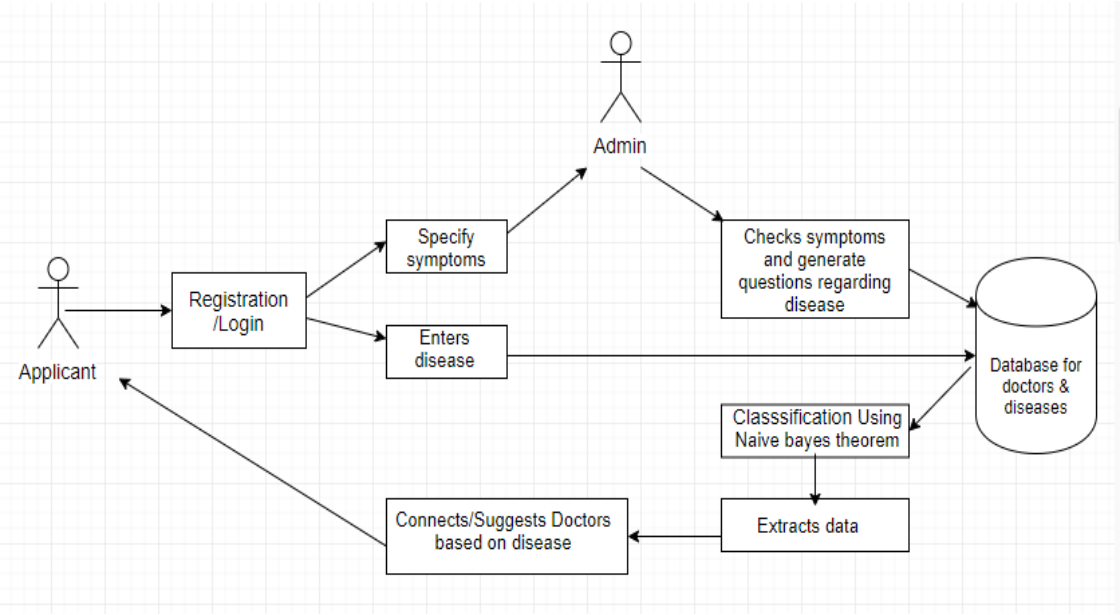
- **Admin Login**: Admin can login to the system using his Id and password.
- **Add Disease**: Admin can add disease ,details of the disease along with symptoms and type.
- **View Patient details:** Admin can view various patient details who had accessed the system.
- **View applicant's Feedback:** Admin can view user's feedback.
- **Location Tracker:** There will be a location tracker which will track Patient's location (from where he accessed the system).

Patient Module:

- **Patient login:** Patient can login to the system using his Id and password.
- **Patient registration:** If patient is a new user,then system will ask him to enter his personal details. After registration,he will have a user id and password through which he can login to the system.
- **My details:** Patient can view his personal details.
- **Edit patient record:** Patient can edit his personal details.
- **Disease prediction:** Patient will specify the symptoms caused due to his illness. System asks certain questions to the applicant regarding his illness and predicts the disease based on the symptoms specified by the patient.System will also suggest doctors based on the disease.
- **Search Doctor with known disease:** Patient can search for doctor by specifying Disease name/ type.
- **My Stuff:** Patient has the facility to manage their reports.
- **Feedback:** Patient will give feedback this will be reported to the admin.

ISSN:0971-1260
Vol-22-Issue-10-November-2019

Architecture of the system



Conclusion

In this proposed system, we develop a system to mine or extract the data which is very important because we are getting valuable information which is not available easily and all this information are real time information. Finally, we want to tell you that this is a fully developed unique system which will be helpful for us. Hope this system will be very demandable in coming future. This system involves fundamental parts, for example, basic login, enter symptoms in the system and recommend medications, proposes an adjacent specialist. It takes the contribution of different manifestations from the patient, does the examination of the entered symptoms and gives fitting sickness expectation.

References

1. Palaniappan S et al. Dept. of IT, Heart disease prediction using techniques of data mining, IEEE April 4.

2. “Survey on Data Mining Algorithms in Disease Prediction”, Kirubha Vetal. IJCTT, 2016

3.Holzinger,Ziefle,Martina,”Smart health”,2015

4.Dean,Jared,”Data mining and Machine Learning”,2014.