

ABSTRACT

Here we propose a system that allows users to get instant guidance on their health issues through an intelligent health care system online. The system allows user to share their symptoms, issues and also he/she may directly search doctors for their medical issue with their requirements. The system has many features suggesting doctors as per the disease and system lists various expert Doctors available. System then processes user's symptoms to check for various illnesses that could be associated with it. In this we are going to implement a system which is related to smart health prediction in order to reduce the time of a user. The main objective of developing this project is to provide a proper medical guidance to the patient for their health issues by providing accurate results . 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. We have 3 modules; they are patient, doctor, and admin modules. The primary step is, the user (patient/doctor) needs to register if he is new to the application or else he can directly login with his credentials. The admin authenticates the credentials and allows the user to access. There is a facility that user can upload records and there is a feedback system available to the user which is directly visible by the admin. So, the doctor can study the patient and give a proper treatment.

Keywords: Disease prediction, Doctor, Patient, Symptoms, Feedback, Appointment, Admin.

CONTENTS

TITLE	Page No.
ABSTRACT	i
LIST OF FIGURES	ii
LIST OF TABLES	iii
1. INTRODUCTION	
1.1 Introduction	01
1.2 Motivation of work	04
1.3 Problem Statement	05
2. LITERATURESURVEY	
2.1 Data mining and its concepts	06
2.2 Basic Data mining Tasks	08
2.3 Steps for Pre-processing	09
2.4 Evaluation Measures	09
2.5 Multi disease prediction using data mining Techniques	10
2.6 Existing system	10
3. INTRODUCTION	
3.1 Proposed system	11
3.1.1 Basic System Architecture	12
3.2 Algorithm Illustration	12
3.3 Features Available in the system	15
3.3.1 Algorithms	15
3.3.2 Connecting Doctors	17
3.4 Modules	16
3.5 Performance Measures	18
4. DESIGN	
4.1 UML Diagrams	21
4.1.1 Use case Diagrams	22
4.1.2 Class Diagram	24
4.1.3 Sequence Diagrams	25

5. EXPERIMENT ANALYSIS AND RESULTS	
5.1 System Configuration	27
5.1.1 Software Requirements	27
5.1.2 Hardware Requirements	28
5.2 Sample code	29
5.3 Screenshots	39
5.4 Experimental Analysis/Testing	48
6. CONCLUSION AND FUTUREWORK	55
References	56
Appendix	57

1.INTRODUCTION

1.1.Introduction

The health industry has been growing a lot from past few years .This technique has gained a lot of importance in medical areas. It has been calculated that a care hospital may generate five terabytes of data in the year. In our day to day life we have lot of other problems to deal with and we neglect our health problems. So in order to overcome such problem we have designed user friendly website which helps users to get diagnosed from their residence at any time. We also provide an option for booking an appointment with the doctor to discuss health related problems and get diagnosed properly. Data mining is the process of discovering anomalies, patterns and correlations within Large data set using sophisticated mathematical algorithms to predict outcomes. Using techniques, you can use the information to increase revenues, cut costs, improve customer relationships, and reduce risks and more.

The task of actual data mining is to automatically analyze the large quantities of data to extract previously unknown, interesting patterns such as groups of data records unusual records (anomaly detection), and dependencies.[1] In simple words, data mining is defined as a process used to extract usable data from a larger set of any raw data. It implies analyzing data patterns in large batches of data using one or more software. Data mining has applications in multiple fields, like science and research. As an application of data mining, businesses can learn more about their customers and develop more effective strategies related to various business functions and in turn leverage resources in a more optimal and insightful manner. This helps businesses be closer to their objective and make better decisions. Data mining involves effective data collection and warehousing as well as computer processing. For segmenting the data and evaluating the probability of future events, data mining uses sophisticated mathematical algorithms. Data mining is also known as Knowledge Discovery in Data (KDD). The term "data mining is in fact an inapplicable title, because the goal is the extraction of patterns and knowledge from large amounts of data, not the extraction (mining) of data itself. Many classification techniques are available in data mining. There are decision trees, K nearest neighbor, Bayesian classification, artificial neural networks (ANN), support vector machine (SVM) and so on.

Machine Learning :

Machine Learning is the field of study that gives computers the capability to learn without being explicitly programmed. ML is one of the most exciting technologies that one would have ever come across. As it is evident from the name, it gives the computer that which makes it more similar to humans: *The ability to learn*. Machine learning is actively being used today, perhaps in many more places than one would expect.

The Learning for a machine :

A machine is said to be learning from **past Experiences** (data feed in) with respect to some class of **Tasks**, if it's **Performance** in a given Task improves with the Experience.

Types of Learning :

There are 2 types of learning, They are :

- Supervised learning
- Unsupervised learning

Supervised Learning :

Supervised learning is a learning in which we teach or train the machine using data which is well labeled that means some data is already tagged with the correct answer. After that, the machine is provided with a new set of examples (data) so that supervised learning algorithm analyses the training data (set of training examples) and produces a correct outcome from labeled data.

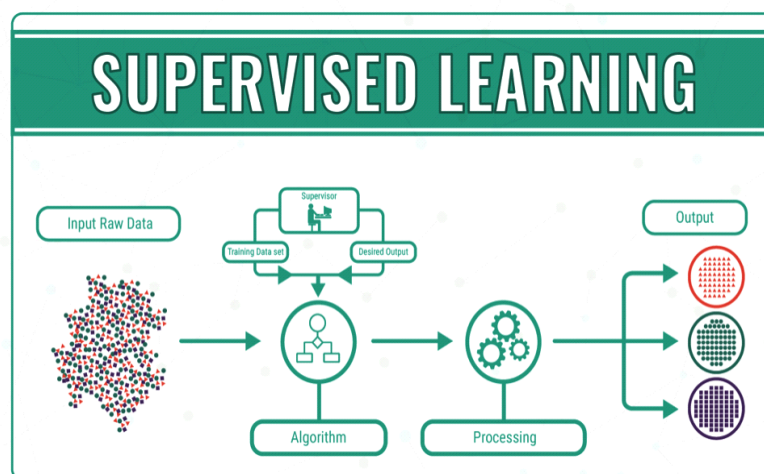


Fig:1.Supervised Learning

Supervised learning classified into two categories of algorithms:

- **Classification:** A classification problem is when the output variable is a category, such as “Red” or “blue” or “disease” and “no disease”.
- **Regression:** A regression problem is when the output variable is a real value, such as “dollars” or “weight”.

Unsupervised Learning

Unsupervised learning is the training of machine using information that is neither classified nor labeled and allowing the algorithm to act on that information without guidance. Here the task of machine is to group unsorted information according to similarities, patterns and differences without any prior training of data. Unlike supervised learning, no teacher is provided that means no training will be given to the machine. Therefore machine is restricted to find the hidden structure in unlabeled data by its-self.

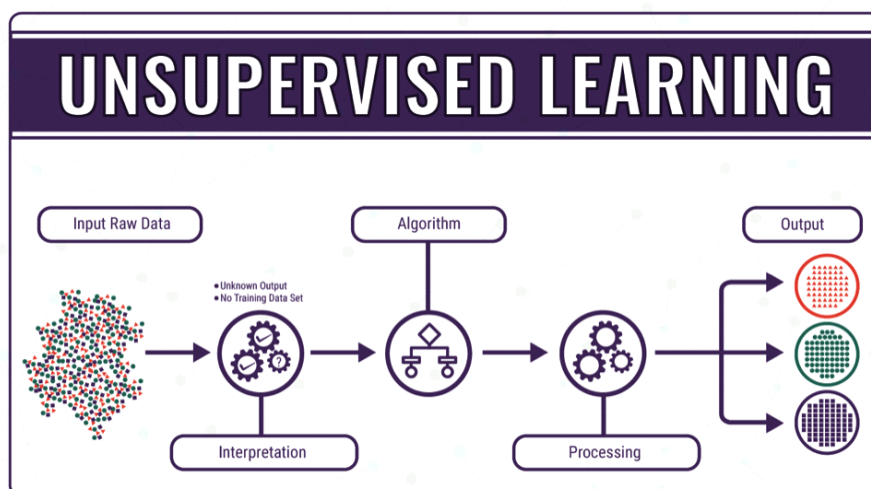


Fig:2.Unsupervised Learning

Unsupervised learning classified into two categories of algorithms:

- **Clustering:** A clustering problem is where you want to discover the inherent groupings in the data, such as grouping customers by purchasing behavior.
- **Association:** An association rule learning problem is where you want to discover rules that describe large portions of your data, such as people that buy X also tend to buy Y.

Naive Bayes Classifiers:

Naive Bayes classifiers are a collection of classification algorithms based on **Bayes' Theorem**. Naive Bayes algorithm is not a single algorithm, but it is a family of algorithm based on a common principle. This principle states that the value of each feature is independent of values of other features of all Naive Bayes classifiers. There are many probability models, out of which for some types the naive Bayes algorithm is efficiently trained in supervised learning training. 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.

Where **A** and **B** are events and **P (B)** is true.

Basically, We are trying to find probability of event A; given the event B is true.

Event B is also termed as **evidence**.

P (A) is the **prior** 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.

$P(A/B)=[P(B/A)*P(A)]/P(B)$ Eq(1.1)

1.2.Motivation of work

Affecting from diseases are becoming more common in our daily lives, major challenge facing healthcare organizations is the provision of quality services at affordable costs and diagnosing patients correctly and administering treatments that are effective. Poor clinical decisions can lead to disastrous consequences which are therefore unacceptable. Hospitals must also minimize the cost of clinical tests. They can achieve these results by employing appropriate computer-based information and decision support systems.

Most hospitals today employ some sort of hospital information systems to manage their healthcare or patient data. These systems typically generate huge amounts of data which take the form of numbers, text, cars and images and an important question "How can we turn data into useful information that can enable healthcare practitioners to make intelligent clinical decision This is the main motivation this data is rarely used to support clinical decision making This raises an important question: "How can we turn data into useful information that can enable healthcare practitioners to make intelligent clinical decisions?" This is the main motivation for this research.

1.3. Problem Statement

The main problem here is that, more time is taken to predict the disease corresponding to the problems. This has to be overcome. A general solution has to be developed which will provide facilities to result the health issues in a faster and more efficient way and user have to search best doctor which consumes a lot of time and then book appointment, so there is a lot of manual work for user.

2.LITERATURE SURVEY

2.1. Data Mining and its concepts

Divya Jain et al presents a review of the implementation of Apriori Algorithm on datasets using machine learning tool Weka. Ruijuan Hu states the details of the idea on two-step frequent data items using Apriori algorithms and Association Rules. This mentions a new improvised method called Improved Apriori Algorithm to eliminate cons of Apriori algorithm. Gitanjali J, et.al proposed study of huge datasets from various angles and obtaining gist of useful information.

Simovici proposed that association rules represent knowledge in data sets as results and are directly related to calculation of frequent item sets. Mohammed Abdul Khaleel states data mining as a concept data and knowledge. that studies large amount of extract patterns that can be converted to useful also to provide these methods are useful in detecting diseases and providing proper remedy for the same. Krishnaiah et.al. aims to calculate various methods of data mining in applications to discussion about medical. Data mining techniques can improve various angles of clinical predictions. Dan A. a detailed develop decisions and In this paper, we set out to identify efficient algorithm for mining results. We can create versatile applications for medicine sector so as to fulfill by using all these predictive analytics and data mining techniques.

1. This tells how Naive Bayes algorithm is used to find frequent data items and compares them with the existing algorithms.
2. Electronic health records and other historical medical data can prove miracles if used for a right purpose.
3. Huge amounts of complex data generated by health care sector include details about diseases, patients, diagnosis methods, electronic patient's details hospitals resources.

The widespread of the databases and the explosive growth in their sizes, organization are faced with the problem of information overload. The problem of effectively utilizing these massive volumes of data is becoming a sior problem for all enterprises. OLAP tools are relevant for decision support systems but they lack exploratory characteristics of querying. Data mining attempts to source out patterns and trends in the data and inference rules from these patterns.Data Mining, the extraction of hidden predictive information from large databases, is a powerful new technology with great potential to analyze information in the data warehouse. Data mining is a non-trivial process of identifying valid, novel, and potentially useful and ultimately understanding patterns in data.Data mining processes have required an integration of techniques from multiple disciplines such as, statistics, machine learning, database technology, pattern recognition, neural networks, information retrieval and spatial data analysis.

Data mining algorithms can be characterized on the basis of

1. **Model:** The purpose of the algorithm to fit a model to the data
2. **Preference:** The criteria that is being used.
3. **Search:** the search techniques that is being used.

The data mining models are of two types:

Predictive Model: This makes a prediction about the values of the data using known results found from different data. Predictive data modeling may be made based on the use of historical data. It includes Classification, Regression, Time Series Analysis and Prediction.

Descriptive Model: This identifies patterns or relationships in data. Unlike the Predictive model a descriptive model serves a way to explore the properties of data examined, not to predict new properties. It includes Clustering, Summarization, Association Rules and sequence discovery.