

Data Analysis Using Dashboard

¹Prof. Pragati Pejlekar, ²Samrutti Patil, ³Archana Patil

¹Professor, ^{2,3}Students

Department of Information Technology Engineering
Saraswati College of Engineering, Kharghar, Navi Mumbai, India

Abstract: Nowadays Healthcare is one of the biggest sectors in India. The biomedical sector generating enormous amount of data each year. This data generated by the Healthcare sector can be used to get meaningful insight about patient, disease which help in improve healthcare system. Accurate study of medical dataset benefits early disease prediction. When the quality of medical data is incomplete the exactness of study is reduced. Hence, we need data completeness. In proposed system it provides machine learning algorithm for prediction of various diseases [1]. It is used to predict Healthcare disease datasets at an early stage. One of our main objectives is to predict disease data. Currently many hearts, diabetics, blood pressure rate diseases are increase throughout world. To group and predict symptoms in medical data, various data mining techniques are used. In this system R Tool and RStudio is used, the predictive algorithms used are Naïve Bayes and KNN. By using this algorithm, we make a model combining Techniques/method in one in order to increase the performance and accuracy. We calculated the accuracy using Naïve Bayes and KNN algorithms. The algorithm with high accuracy is used for prediction. The exactness of our proposed algorithm will be good and high [1].

Keywords: - Healthcare, disease, data, algorithm, Naïve Bayes, KNN, Prediction, RStudio.

I. INTRODUCTION

Data mining is the technique rapidly increasing in the field of healthcare community. Health data mining aims to discover hidden data and knowledge about patient. The health care prediction is important issue of improving healthcare process. Patient health can be affected due to many factors. The early prediction of disease may improve patient health. To solve these problems, it sees the structured and unstructured data in healthcare field to assess the risk of disease. The system uses KNN and Naïve Bayes algorithm to generate the pattern and causes of disease. It clearly shows the diseases and sub diseases [1]. Naive Bayes and KNN Algorithm are also used for prediction of disease to get better accuracy. The aim is to achieve good accuracy from the system.

Data analysis: it is the process of cleaning, transforming data to gain useful information.

Predictive analytics: the retrieving information from existing data sets in order to find patterns and predict future outcomes and trends. The disease which can occur in future depending on current symptoms is predicted and solution to avoid that disease is provided.

II. SYSTEM DETAILS

After studying the existing system, we derived the dashboard which will predict disease and sub disease. The system will take data as input and derive the output. After input data system will preprocess data for clearing the data. The dataset is then divided into two part one for training and other for testing, basis of given input data file system will train itself using machine learning algorithm [7]. Here for prediction we use naïve Bayes and KNN algorithm. The software used for machine learning in RStudio and R Tool. R language is used by RStudio for machine learning, R has all the types of machine learning library and function inbuilt which make it very simple and flexible to use.

Using the train data dashboard will display the accuracy prediction using KNN and Naïve Bayes algorithms. The accuracy of both the two algorithm is compared and the algorithm with best accuracy is used for prediction of disease and sub disease.

In general, the existing systems have the following features-

1. Future prediction of disease.
2. Able to find disease and sub disease.
3. Low cost.
4. Time efficient process.

III. PROPOSED SYSTEM

The system which is build will be at affordable cost. KNN Algorithm and Naïve Bayes Algorithm are used to predict diseases as well as all sub types of diseases. The proposed system will consider both structured and unstructured data. The analysis accuracy is increased by using Machine Learning Algorithm the data of patient is collect and stored in file format for further processing [1]. RStudio is used for processing, cleaning and learning the machine. All type of different data is collected and stored in Data collection. In Data processing, the Data which is stored is cleaned. The data is divided into training and testing sets and the algorithms were applied and proper prediction of data is done [7]. Then system compares the prediction using both algorithms. The result is obtained with high accuracy prediction.

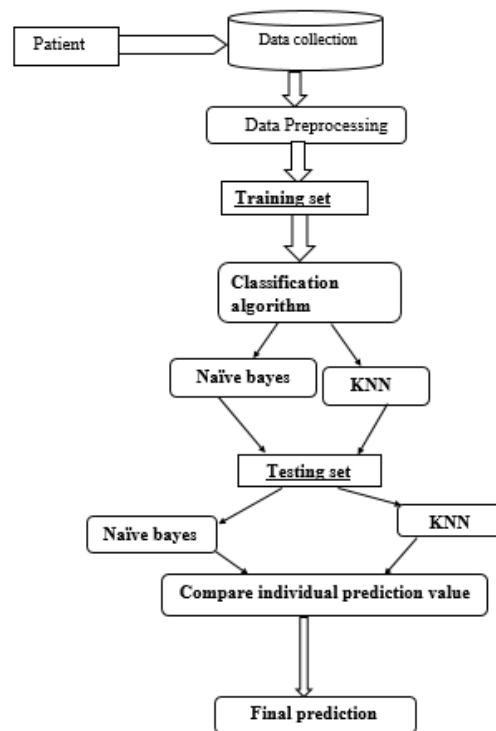


Fig: III.1. Architecture diagram

The system flow given below:

Data pre-processing this step is very important. It will give the accurate data. In these steps the empty values get filled and provide accurate data for further process.

Classification process: here in system there are mainly two classification algorithms are used namely naïve Bayes and KNN to give the proper accuracy.

Prediction: with use of KNN and Naïve Bayes algorithm system predicts disease and sub disease.

Proposed algorithm: -**KNN algorithm:**

K-nearest neighbor is classification algorithm. it is very easy and simple classification algorithm. This algorithm classifies new neighbor based on similarity. The value of neighbor assign is always a positive integer. The algorithm stores training data. On the basis of nearest neighbor, the prediction of data is done.

Steps to compute K-nearest neighbor:

StepI: let K =number of nearest neighbors.

StepII: calculate gap between query-occurrences and all training data samples.

StepIII: find out the distance and determine distance based on K -the minimum distance in system.

StepIV: group all the nearest neighbors in data.

StepV: use simple large part of nearest neighbor's as the prediction value for problem.

Advantages:

1. Trains data fast and efficiently.
2. It is better to inconsistent training data
3. Ease to understanding and implement.

Disadvantages:

1. It is very slow learner.
2. It is sensitive to local structure of data
3. Memory cost is more.
4. It runs very slowly, as it is lazy learner.

Naïve Bayes algorithm:

Naïve Bayes is the best classification technique based on Bayesian theorem. It works precisely and efficiently when provided large data. The Bayesian user can add row data at run time of system and it will be able to give the better prediction. This algorithm is able to calculate more output by given inputs.

Advantages:

1. Takes less time for training data.
2. It provides better performance than other classification algorithm.
3. Clearing the features will improve the performance of algorithm.
4. It provides better performance than other classification algorithm.

Disadvantages:

1. These algorithms require large dimension data to give good accuracy, prediction. Less accurate.

IV.RESULT

1. System will generate accuracy using KNN and Naïve Bayes algorithm.
2. The algorithm which gives good accuracy is used for disease prediction.
3. The aim is to achieve good accuracy from the system.

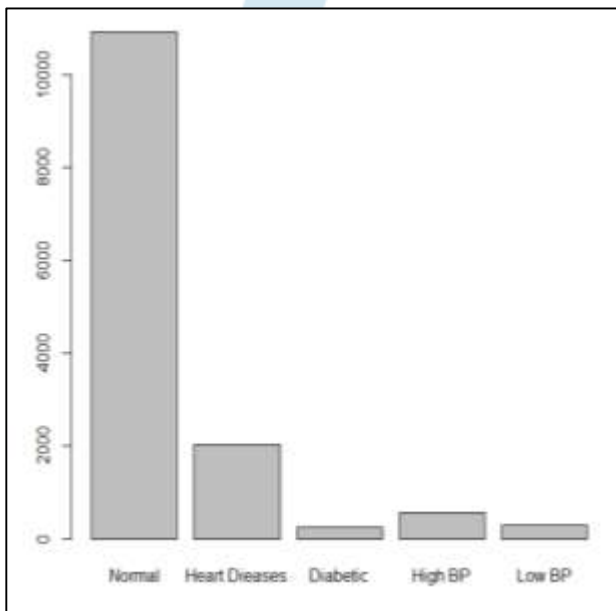


Fig IV.1. Analysis

Fig IV.2: Prediction of Diabetic using Naive Bayes.

The screenshot shows a web form with various input fields for patient data. The 'Result' field on the right side of the form displays the text 'Heart Diseases'.

Fig IV.3: Prediction of Heart disease

The screenshot shows a web form with various input fields for patient data. The 'Result' field on the right side of the form displays the text 'High BP'.

Fig IV.3: Prediction of sub disease

V.ADVANTAGES

1. The system uses KNN and Naïve Bayes easy algorithm to generate the pattern and causes of disease.
2. It clearly shows the diseases and sub diseases.
3. Naïve Bayes and KNN Algorithm are also used for prediction of disease to get better accuracy.
4. The prediction made should be proper and solution to problem is provided

VI.FUTURE SCOPE

1. This System will be used for real life Use of machine learning algorithms and get good accuracy prediction for disease.
2. In much application we can use this system to future prediction of results.
3. Examples for weather forecasting, temperature prediction, business risks and profit prediction.

VII.CONCLUSION

In this project we have presented a way in which machine leaning techniques can be applied to large sets of data. The prediction achieved using decision KNN algorithm and naive Bayes algorithm. Both are prediction algorithms. The result which should be obtain be with high accurate and the processing speed of data, the training, learning and prediction of future using prediction algorithm all these should done very quickly without any delay. Machine learning KNN algorithm and Naïve Bayes algorithm by using structured and unstructured data from hospital.

REFERENCES

- [1] Vinitha S:DISEASE PREDICTION USING MACHINE LEARNING OVER BIG DATA.
- [2] "M. Chen, S. Mao and Y. Liu. Big data: A survey".
- [3] Hughes G. How big is 'Big Data' in healthcare?
- [4] Choong Ho Lee and Hyung-Jin Yoon Medical big data: promise and challenges.
- [5] <https://www.kaggle.com/sulianova/cardiovascular-disease-dataset>
- [6] Big data is the future of Healthcare.
- [7] Exploratory Analysis in Time-Varying Data Sets: a Healthcare Network Application Narine Manukyan, Margaret J. Eppstein, Jeffrey D. Horbar, Kathleen A. Leahy, Michael J. Kenny, Shreya Mukherjee, and Donna M. Rizzo
- [8] A Rishika Reddy Predictive Big Data Analytics in Healthcare.
- [9] Rashmi G Saboji:A Scalable Solution for Heart Disease Prediction using Classification Mining Technique
- [10] Emrana Kabir Hashi:An Expert Clinical Decision Support System to Predict Disease Using Classification Technique.