

A Smart Clinical Decision Support System to Predict diabetes Disease Using Classification Techniques

¹K. Lakshmi, ²D.Iyajaz Ahmed, ³G. Siva Kumar

¹Assistant Professor, Department of Computer Science and Engineering, G. Pullaiah College of Engineering and Technology, Kurnool, , India

^{2,3} B.Tech, Department of Computer Science and Engineering, G. Pullaiah College of Engineering and Technology, Kurnool, , India

ABSTRACT

In the present living scenario a diseases are one of the crucial reason for the increasing death rates. The major problem is inaccurate identification of diseases which give rise to a complex task which needs to be evaluated specifically and efficiently with a desirable automated technology. Different human beings have different skills, so as doctors. Due to the lack of specialized doctors in remote areas the health services are not availed easily.as in our presented system there are several classification techniques available that can be used for clinical decision support system different techniques are used for different diagnosis.As our proposed system uses the Decision Tree and K-Nearest Neighbor (KNN) Algorithms as supervised classification model for diabetes disease because it is a great threat to human life worldwide finally our proposed system will reduce the time and cost of diagnoses.

Keywords : Diabetes, data mining, Decision Tree and K-Nearest Neighbor (KNN) Algorithms.

I. INTRODUCTION

Clinical Decision Support System is a healthcare system, which is intended to assist physicians and other health professionals on decision making tasks. In healthcare sector, the term information mining can mean to analyze the clinical information to predict patient's health status. So discovering interesting pattern from healthcare data, different data mining techniques are applied with statistical analysis, machine learning and database technology. It can be also defined as "A computer system that uses two or more patient data to generate case specific or encounter specific advice." Most CDSS (Clinical Decision Support System) comprises of three parts, the knowledge base, inference engine, and mechanism to communicate. The knowledge base comprises of compiled information that is often, but not always, in the form of if-then rules. The inference engine comprises the expressions for combining the rules or associations in the knowledge base with

actual patient data. Then a communication mechanism is used for bringing the patient data into the system and supplying the output of the system to the user who will make the actual decision [7]

II. System Study

2.1 Existing System:

In our presented system the conventional disease diagnosis system uses the perception and experience of doctor without using the complex clinical data. In medical science, doctor's facilities introduced different data frameworks with a lot of information to manage medical insurance and patient information but unfortunately, data are not mined to discover hidden information for effective decision Clinical test outcomes are regularly made on the basis of doctors' perception and experience rather than on the knowledge enrich data masked in the database and sometimes this procedure prompts inadvertent

predispositions, doctor's expertise may not be capable to diagnose it accurately which affects the disease diagnosis system [2].

2.2 Proposed system:

As our proposed system is created to predict different diseases. As our proposed system uses the Decision Tree and K-Nearest Neighbor (KNN) Algorithms as supervised classification model for diabetes disease because it is a great threat to human life worldwide finally our proposed system will reduce the time and cost of diagnoses.[5]

1. Attributes for diabetic

Datasets are comprises of 11 attributes which are as follows:

- 1) Age (years)
- 2) Sex
- 3) Body mass index
- 4) Blood Pressure (mm Hg)
- 5) Plasma Glucose Concentration (Glucose tolerance test)
- 6) Triceps Skin fold
- 7) 2-Hour serum insulin
- 8) Diabetes Pedigree function
- 9) Cholesterol Level
- 10) Weight (kg)
- 11) Class variable (0 or 1)

2. Algorithm Used

Decision Tree:

Decision tree works supports with both nominal and numerical features. It can be robust towards the noise and inconsistent values. Decision tree follows the top down approach and categorizes the entire trained dataset by partitioning them from top most nodes to the class node. Each node represents the test attribute of the instance and the nodes related to one of the distinct possible values for that feature attribute. A decision tree can easily transform the given set of instances into meaningful patterns from the top node to the attack class node level by level [6]

4.2 The k-means algorithm: is well known for its efficiency in clustering large data sets. However,

working only on numeric values prohibits it from being used to cluster real world data containing categorical values. For prediction the majority of practical machine learning uses supervised learning. Supervised learning is where you have input variables (x) and an output variable (Y) and you use an algorithm to learn the mapping function from the input to the output [1].

$$Y = f(X)$$

The goal is to approximate the mapping function so well that when you have new input data (x) that you can predict the output variables (Y) for that data.

It is called supervised learning because the process of algorithm learning from the training dataset, now our proposed algorithms are Decision Tree and K-Nearest Neighbor (KNN) Algorithms as supervised classification model for diabetes disease.

Proposed System Architecture

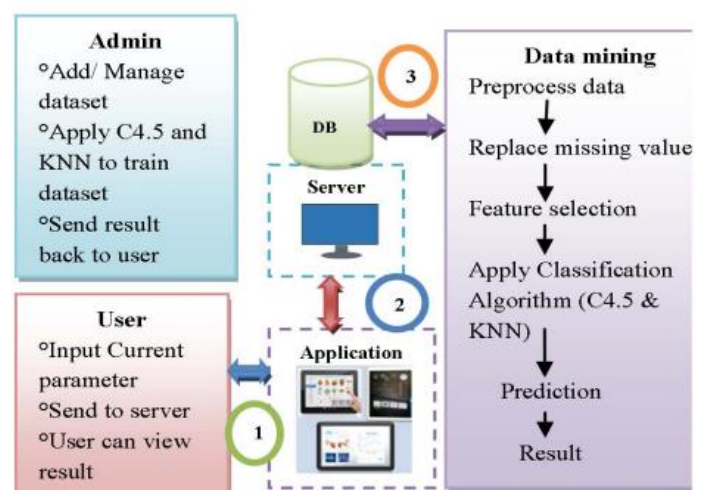


Figure 1 : Proposed System Scenario

Modules:

Admin: On the server, admin can load dataset of different diseases and apply different data mining algorithms to train dataset. Requested user inputs are collected and processed on server to predict the diagnosis result.

User: Through the proposed application user (doctor, patient, physician etc.) can input the attribute values of disease and send it to the server with the help of internet. After applying the data mining approach the predicted result can be viewed on the user GUI [7].

Data mining: For analyzing healthcare data, major steps of data mining approaches like preprocess data, replace missing values, feature selection, machine learning and make decision are applied on train dataset. On the server different algorithms[3].

Diabetes data using various data mining techniques which involved,] Naive Bayes, J48, Neural networks, Decision trees, kNN, Fuzzy logic and Genetic Algorithms based on accuracy and time was analyzed by VelidePhani Kumar et al . They found that out of various data mining techniques which were employed to analyze the diabetes data.

As per our proposed algorithms kNN has got the highest accuracy of 95% and proved that in medical data such as diabetes disease.

KNN is a method which is used for classifying objects based on closest training examples in the feature space. KNN is the most basic type of instance-based learning or lazy learning. It assumes all instances are points in n-dimensional space. A distance measure is needed to determine the “closeness” of instances. KNN classifies an instance by finding its nearest neighbors and picking the most popular class among the neighbors [8].

Features of KNN

- a) All instances of the data correspond to the points in an n-dimensional Euclidean space
- b) Classification is delayed till a new instance arrives
- c) In KNN, the Classification is done by comparing feature vectors of the different points in a space region.
- d) The target function may be discrete or realvalued.

Decision tree works supports with both nominal and numerical features. It can be robust towards the noise and inconsistent values. Decision tree follows the top down approach and categorizes the entire trained dataset by partitioning them from top most node to the class node. Each node represents the test attribute of the instance and the nodes related to one of the distinct possible values for that feature attribute. A decision tree can easily transform the given set of instances into meaningful patterns from the top node to the attack class node level by level.

III. Conclusion

In this paper we presented data mining algorithms clustering and classification algorithms like Decision Tree and K-Nearest Neighbor (KNN) Algorithms as supervised classification model for diabetes disease prediction. Because it is a great threat to human life worldwide finally our proposed system will reduce the time and cost of diagnoses.

IV. REFERENCES

- [1]. R. Fang, S. Pouyanfar, Y. Yang, S. Chen and S. Iyengar, "Computational health informatics in the big data age: a survey," *ACM Comput. Surv.*, New York, Vol. 49, pp. 12-47, June 2016.
- [2]. P. K. Anooj, "Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules," *J. of King Saud Uni. Comput. and Inform. Sci.*, ELSEVIER, Vol. 24, pp. 27-40, 2012.
- [3]. Purushottam, K.Saxena and R. Sharma, "Efficient Heart Disease Prediction System," *Proced. Comput. Sci.*, ELSEVIER, Vol. 85, pp. 962 – 969, 2016.
- [4]. P. Agrawal and A. Dewangan, "A brief survey on the techniques used for the diagnosis of diabetes-mellitus," *Int. Res. J. of Eng. and Tech.* IRJET, Vol. 02, pp. 1039-1043, June-2015.
- [5]. Elma Kolçe (Çela), Neki Frasher, "The Use of Heuristics in Decision Tree Learning Optimization", *International Journal of Computer Engineering in Research Trends.*, vol.1, no.3, pp. 127-130, 2014.
- [6]. Havinash P.H,Jeril Johnson N, Glen Thomas,Emily Stephen P," Mining Opinion Features in Customer Reviews", *International Journal of Computer Engineering in Research Trends.*, vol.3, no.9, pp. 535-539, 2016.
- [7]. HIMABINDU MUTLURI, Mrs P.SUJATHA," Challenges in Big Data using Data Mining Techniques.", *International Journal of Computer Engineering in Research Trends.*, vol.2, no.12, pp. 924-930, 2015.
- [8]. Milind M. Sutar, Tanveer I. Bagban," Survey on: Prediction of Rating based on Social Sentiment", *International Journal of Computer Engineering in Research Trends.*, vol.4, no.11, pp. 533-538, 2017.