International Conference on Advancements in Computing Technologies - ICACT 2018 ISSN: 2454-4248
Volume: 4 Issue: 2 71 – 74

Predictive Analysis of Diabetes Using Bayesian Network and Naive Bayes Techniques

K.Priyadarshini[1], Dr.I.Lakshmi[2]

P.G. Student, Department of Computer Science,Stella Maris College, Chennai, India[1]

Assistant Professor, Department of Computer Science,Stella Maris College, Chennai, India[2]

Abstract: This paper contributes in predicting diabetes by invoking data mining techniques. The detection of knowledge from medical datasets is significant in order to produce operative medical diagnosis. The intention of data mining is to express knowledge from the information gathered in dataset and promote straightforward and transparent interpretation of models. Diabetes is one of the longstanding disease and a vital public health trouble worldwide. And also there is no particular age group to get affected by diabetes. It may occur from a new born baby to any age group of persons. Employing data mining procedure to relief people to predict diabetes has acquire considerable admiration. This paper centralizes on the prediction of diabetes for the dissimilar age groups. Bayesian Network and naive Bayes was intended in order to predict the person's diabetic possibility. In this project Weka tool is used for the assessment and analysis of the dataset. Diabetes is one of the prime universal ailment under which the formulation and utilization of insulin is dispersed everywhere and the body which accordingly leads to the accumulation of glucose level in the blood. The dataset which is used in this project is Pima Indian Diabetes dataset, which gathers the information of persons with and without diabetes

***Keywords**: Diabetes,Data Mining,Prediction,BayesianNetwork,Naive Bayes.*

*****

I. INTRODUCTION

Data mining is usually characterized as the procedure of detecting, interrelationships, designs, associations through analytics by a vast volume of data collected in depositories, medical databases and data warehouses. In the area of medical sector, health informatics is a fastest evolving field with the help of emerging trends in information technology. It also plays a vital role for executing the data mining techniques in diverse sectors like banking, education, fraud detection etc. Though there are various techniques for the prediction of various diseases like heart, kidney, lungs, cancer etc. But there are no enough techniques for the prediction of diabetes disease. Prediction of diabetes is an emerging and one of the fastest growing technology in the field of medical industry. Bayesian networks are believed as supportive techniques for the identification of many diseases. They are actually a probable (most likely) representation which have been determined convenient in presenting composite systems and exhibiting the correlation between variables in a graphic way. This Research paper concentrates on prediction of the diabetes with the existing data mining techniques and also which algorithm can shows the best accuracy among them. Normally a person is assessed to be the victims of diabetes, when blood sugar measures are extended the usual level. Diabetes is considered as one of the metabolic disorder in which a person has high-rise blood glucose level. This high blood sugar generates the indications of continual urination, improved thirst and boosted hunger. It happens when the body does not capable of producing enough insulin. Insulin which is secreted by pancreas, is one of the most important hormones in the human body, which is essential to maintain the level of glucose. Diabetes can be controlled with the help of insulin injection, a healthy diet and a regular exercise. Unprocessed diabetes can lead too many complications. Crucial long-term problems include heart disease, kidney failure, and damage to the eyes and nervous systems. The ultimate aim of the diabetes prediction is to generate awareness among public and they can know the disease well in advance, as it rapidly the saves the human lives. We can use Bayesian network and naive Bayes on the data set of the persons in order to predict whether the person is diabetic or non-diabetic and also the accuracy level. The important three types of diabetes are discussed below:

A. Type 1

Type 1 results from the body's omission to generate insulin. It arises when the pancreas is not qualified of producing insulin at all. Insulin is a hormone which is promoted by the pancreas. Type 1 diabetes can happen at any age. It will occur most frequently among children's and young peoples.

B. Type 2

Type 2 outcomes from the insulin resistance, a state in which cells fails to utilize insulin satisfactorily needs for the body. Sometimes also with an entire insulin insufficiency. It appears when the proportion of insulin is not adequate for the body needs. Due to family genetics, old age, obesity also stretches the possibility of gaining type 2 diabetes. It mostly occurs at the age of 40

C. Gestational diabetes

It is the third main type of diabetes that prone to exist extremely with pregnant women's due to the excessive

raised blood glucose levels as the pancreas cannot initiate enough amount of insulin.

D. Pregestational Diabetes

Presentational diabetes arises when the insulin-dependent diabetes in a person for previously becoming pregnant.

II. LITERATURE REVIEW

The objective of the research paper," Predicting Diabetes by consequence the various Data Mining Classification Techniques" describes the various Data Mining Classification Techniques. There are many classification techniques used in this paper for predicting diabetes [2]

The Research paper, "Disease Prediction in Data Mining Technique"– A Survey. The disease prediction plays an important role in data mining. This paper analyses about various diseases like Heart disease prediction, Breast cancer prediction, Diabetes by using many techniques like Classification, Clustering, Decision Tree, Naive Bayes methods in order to predict the diabetes disease. This paper also tells about predictive and descriptive type about the data. Prediction involves some fields in the data set to predict the values of other variables. On the other hand Description focuses on finding patterns of the data that can be interpreted by humans. The different algorithm of data mining are used in the field of medical prediction are discussed in this paper [1]

The Research paper, "Analysis of various Data Mining Techniques to Predict Diabetes Mellitus", concentrates about overall population affected by diabetes worldwide. This paper also predicts about the overall population affected by diabetes will also double the rate of diabetes of the population by the upcoming years. This Paper aims about the early prediction of the diabetes will save the life of the human. The paper analyses about the three types of diabetes and their causes. It also uses the prediction, classification technique this provides the higher accuracy for the disease prediction [5]

The research paper, "Review on Prediction of Diabetes using Data Mining Technique", elaborates about detailed review of existing data mining methods used for prediction of diabetes. It also gives about the types of diabetes disease Type1,type2,and type3.The aim of the diabetes is to predict the diabetes with the help of Data mining methods such as the K-NearestNeighbor Algorithm, Bayesian Classifier, Naive Bayesian Classifier, Bayesian Network, all the methods are used for prediction of diabetes. This paper also mentions about the effects of diabetes on patients [7]

The research paper, "survey on Naive Bayes Algorithm for Diabetes Data Set Problems", explores about various Data mining algorithm approaches of data mining that have been utilized for diabetic disease prediction. In this paper Classification and Naive Bayes is one of the most used algorithm for the prediction of disease [3]

The research paper," Prediction of Diabetes using Bayesian network", describes about the Bayesian network, classification. By using this effective algorithm methods diabetes prediction can be done [4]

The research paper, "prediction of diabetes using probability approach. This paper describes about the Disease Prediction, Bayesian network algorithm, naïve Bayes in data mining by using probability theory. The main aim of this paper is to find out best model from different algorithm that can be used to predict diabetes by applying on the data set of the patients [6]

III. PROPOSED METHODOLOGY

A. Bayesian network

Bayesian Networks are specialized as joint conditional probability distributions. They are also most popularly known as Belief Networks or Probabilistic Networks. It is a graphical characterization of the probabilistic correlation between a categories of variables. This is one of the efficient algorithm which is mainly used for the prediction of disease in data mining. BN is presented in the format of a directed graph whose nodes illustrates the attributes and edges symbolizes the connection among them. Each node is a directly regulated acyclic graph officially presents a random variable. Bayesian network can also be useful to learn causal relationships and hence can be used to achieve considerable knowledge about a problem specified domain and to predict the seriousness of interference. It takes a specified sets of inputs and delivers an efficient output for the predictable model. Moreover Bayesian networks are probabilistic since they are construct from the source of probability distributions and also majorly uses the law of probability for prediction. It is also predominantly useful for abnormal innovation for predictable and identification of selection invention under unreliability and also for time series prediction.

B. Naive Bayes

Naive Bayes is an uncomplicated technique that can rely on Bayes theorem with hypothesis among predictors for individualistic assumptions. It is trouble-free and rapid to forecast class of test data set. Traditionally it appeal for a large number of data sets. It works for probability theory to classify data. A Naive Bayesian pattern is familiar to establish and it has no completeness of iterative boundaries. Naive Bayes is a simple and dynamic algorithm which is more or less used for the classification task. It execute efficiently in the case of absolute input variables contrast to number variables. Naive Bayes renders considerable outcomes when it can be administered it for data analysis. The Naive Bayes empowers to briskly design frameworks that presents predictive means and also carry a new manner for traversing and interpretation of the data. This performance can be able to apply in predictive analysis when constructing a predictive representation with Naive Bayes. For all this the input attributes must be comparatively independent. Naive Bayes is a strong and powerful predictor. This technique can be useful for very large number of data sets. It represents a set of self-standing variables. Naive Bayes behaves continuously prior and later the minimizing of attributes with same representation of building time. The visualization distributed for naive bayes are very much straightforward to understand. Therefore, it could be used for inventions in predictions for real time model as well.

 ISSN: 2454-4248
71 – 74

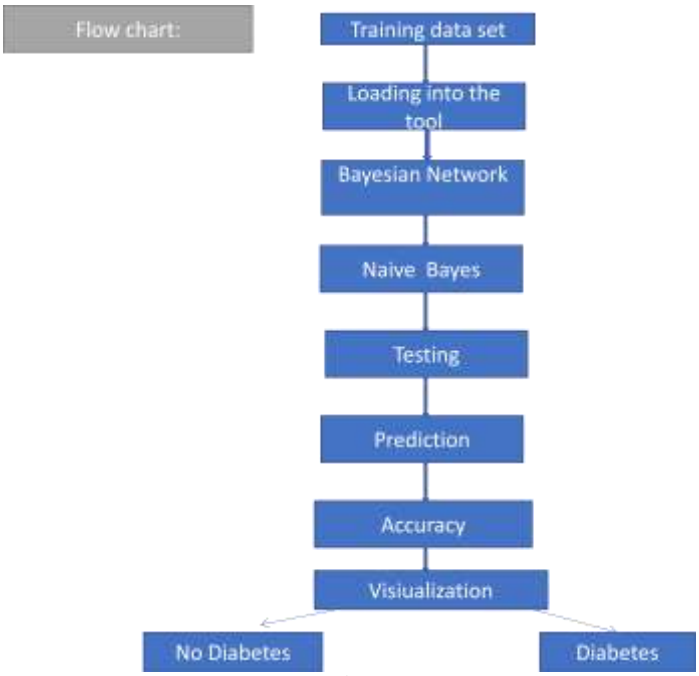

Fig1. The taxonomy of this project work

IV. DATA SET DESCRIPTION

The dataset consists of nine attributes and 1540 instances are obtained from the National Institute of Diabetes Diseases. The Attributes values are tabulated in the table are as shown:

ATTRIBUTE ID	ATTRIBUTE DESRIPTION
1	Number of times Pregnant
2	Plasma glucose test
3	Blood Pressure
4	Skin old Thickness
5	Insulin Level
6	Body Mass Index
7	Diabetes pedigree function
8	Number of Years (Age of person)
9	Class(string) 0=no ,1=yes

V. PERFORMANCE ANALYSIS

For performance Analysis need to be done, the training data set is said to be loaded into the performing tool. Training datasets are extracted as the inputs to be given and it describes attributes of diabetes .The attributes that are described, using this we can observe the connection to every attributes in detailed for diabetes prediction. In order to discover stated individuals in dataset whether will be diabetic, non-diabetic or post diabetic will be accepted on the base of attribute value. The diabetes dataset comprises of with and without diabetic occurrences and nine attributes. By applying the Bayesian network and Naïve bayes algorithms on the training data set we can see the diabetes prediction of the patients in terms of the accuracy level among them. Based on the accuracy level of the algorithm, we can come to the conclusion that which algorithm works way better than the other algorithm. Accuracy is a measure of the percentage of predictions that are accurate.

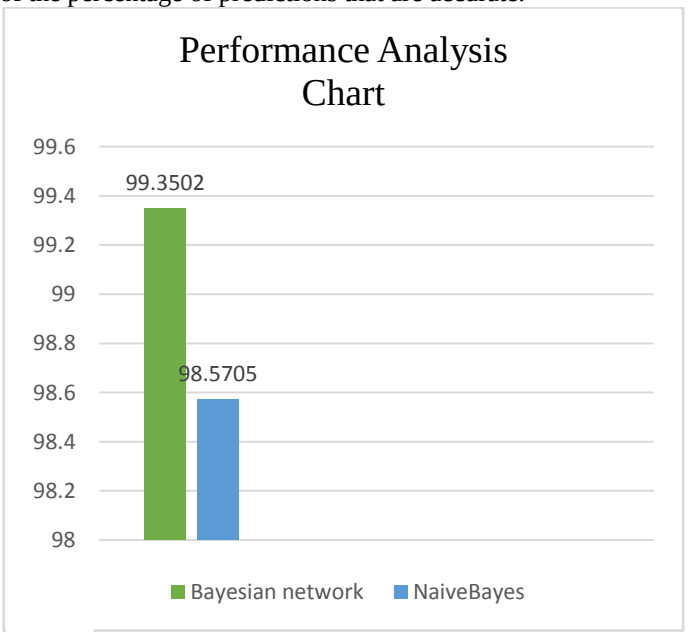

Fig 2. Performance Analysis Chart

VI. EXPERIMENTAL RESULTS

The nature of the work is about predicting whether a person is diabetic or not by applying Bayesian network and naïve Bayes in a data set for the 1540 instances in order to find the best prediction model. Since applying both the algorithms, eventually the accuracy was shown based on the attributes values. The quality of the prediction models is assessed by Accuracy. The Accuracy is the proportion of testing set examples that is correctly categorized by the model. The accuracy level can be varied based on the dataset types and also the number of instances that the data set have. The Accuracy of the model is calculated and compared with other algorithms efficiently. The Experimental Results have been found based on the 1540 instances. The large number of instances, the more number of accuracy was founded. For Bayesian network the accuracy percentage was obtained is 99.3502% for the time seconds of 0 and for Naive Bayes is 98.5705 % for the time seconds of 0.02.

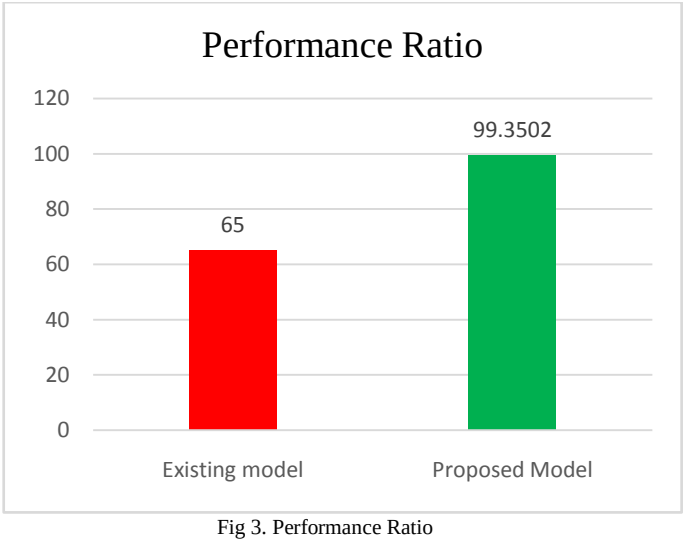

Fig 3. Performance Ratio

VII. CONCLUSION

This project work concentrates about data mining techniques and methods which are used for the early prediction of a diabetes from the medical data set of the patient. Since diabetes is a persistent and severe disease. An advance prediction of the disease will conserve and save the Human life. Hence by applying Data mining approaches and algorithms will helps to predict the diabetes and also generates awareness among public. In this manner data mining process are applied and analyzed in medical data domain in order to predict diabetes and to find out its performances in efficient ways to predict them as well.

REFERENCES

[1] S.VijiyaraniS.Sudha,“ Disease Prediction in Data Mining Technique”– A Survey,International Journal of Computer Applications & Information Technology Vol. II, Issue I, January 2013 (ISSN: 2278-7720)

[2] S.VijiyaraniS.Sudha,“ Disease Prediction in Data Mining Technique”– A Survey,International Journal of Computer Applications & Information Technology Vol. II, Issue I, January 2013 (ISSN: 2278-7720)

[3] S.VijiyaraniS.Sudha,“ Disease Prediction in Data Mining Technique”– A Survey,International Journal of Computer Applications & Information Technology Vol. II, Issue I, January 2013 (ISSN: 2278-7720)

[4] P. Radha , Dr. B. Srinivasan, “Predicting Diabetes by cosequencing the various Data Mining Classification Techniques”,IJISET - International Journal of Innovative Science, Engineering & Technology Vol. 1 Issue 6, August 2014

[5] NileshJagdishVispute, Dinesh Kumar Sahu, Anil Rajput,”ASurvey on naive Bayes Algorithm for Diabetes Data Set Problems”, International journal for research in Applied Science & Engineering Technology (IJRASET),Volume 3 issue XII,December 2015

[6] HalduraiLingaraj, RajmohanDevadass, VidyaGopi, KalirajPalanisamy,” PREDICTION OF DIABETES MELLITUS USING DATA MINING TECHNIQUES”:A REVIEW, Journal of Bioinformatics &Cheminformatics,Feburary 19,2015.

[7] Dr.M.Renuka Devi,J.MariaShyla,”Analysis of various Data Mining Techniques to Predict Diabetes Mellitus”, International Journal of Applied Engineering Research ISSN 0973-4562 Vol 11, Number 1(2016).

[8] IshaVashi, Prof. ShailendraMishra,”A Comparative Study of Classification Algorithms for Disease Prediction in Health Care”, International Journal of Innovative Research in Computer and Communication Engineering, Vol. 4, Issue 9, September 2016.

[9] VrushaliBalpande, RakhiWajgi,” Review on Prediction of Diabetes using Data Mining Technique”, International Journal of Research and Scientific Innovation (IJRSI) |Volume IV, Issue IA, January 2017 | ISSN 2321–2705.

[10] Webresource:http://conexion.itnova.co/wpcontent/uploads/2014/12/DataMiningSQLServer2008.pdf

[11] Mukesh kumari1, Dr.RajanVohra 2,Anshul arora3“Prediction of Diabetes Using Bayesian Network”International Journal of Computer Science and Information Technologies, Vol. 5 (4) , 2014, 5174-5178

[12] T.monikasingh,Rajashekrshastry,”Prediction of diabetes using Probability Approach International Research Journal ofEngineeringand Technology (IRJET)Volume: 04 Issue: 02 | Feb -2017