International Journal of Technical Innovation in Modern Engineering & Science (IJTIMES)
Impact Factor: 5.22 (SJIF-2017), e-ISSN: 2455-2585
Volume 5, Issue 04, April-2019

A REVIEW ON PREDICTION OF DIABETES USING VARIOUS TECHNIQUES IN HEALTHCARE

Miss Vaishali G. Mohature

Department of computer science and engineering SSGMCE, Shegaon, India

Abstract :- ***Now a days, there are many applications are used for searching results on web. In this system we predicting diabetes by applying data mining technique. Data mining is the process of discovering interesting pattern and large amount of data. The main aim of this project is to build a basic decision support system which can determine and exact previously unseen patterns, relation and concepts related with multiple disease data mining is an engineering study of extracting previously undiscovered patterns from a selected set of data. Data set used is Pima Indian diabetes dataset, which collects the information of person with and without diabetes. Data mining is one of the fastest growing field in the health care industries. Using data mining method to aid people to predict diabetes disease has gain major popularity.***

Keywords- Diabetes, Prediction, Naïve Bayes, Decision tree, SVM, Classification, and Dataset.

I.INTRODUCTION

Data mining is one of the most important technique used in the Medicine analysis. The disease prediction plays an important role In data mining. Data mining holds great potential in the health care industry to enable health systems to systematically use data and analytics to identify in efficiencies and best practices that can improve care and reduce cast. Many researchers are conducting experiments for diagnosing the diseases using various classification algorithms of machine learning approaches like SVM, Naive Bayes, Decision Tree, Decision Table etc. In this work, Naive Bayes, SVM, and Decision Tree machine learning classification algorithms are used and evaluated on the PIDD dataset to find the prediction of diabetes in a patient. Experimental performance of all the three algorithms are compared on various measures and achieved good accuracy [5]. Using data mining methods to aid the people to predict diabetes has gain major popularity. Data mining is the process of discovering correlations, patterns or relationships through large amount of data stored in repositories, databases and data warehouse. Many techniques or solutions for data mining and knowledge discovery in databases are very widely provided for classification, association, clustering and regression, search, optimization. Diabetes is the fast growing disease among the youngsters. In diabetes a person generally suffering from high blood sugar. Intensify thirst, Intensify hunger and frequent urination are some of the symptoms caused due to high blood sugar. There are different type of disease predicted in the data mining i.e. sugar, breast cancer, lung cancer, thyroid etc.

Background

In the background we have defined the basic definitions that can be used for diabetes prediction. The discovery of knowledge from medical datasets is important in the order to make medical diagnosis.

Diabetes is a chronic disease that occurs when the human pancreas does not produce enough insulin, or when the body cannot effectively use the insulin it produces, which leads to an increase in blood glucose levels.

There are generally three types of diabetes,

TYPE 1- In this type of diabetes, the pancreatic cells that produce insulin have been destroyed by the defense system of the body. This type can be caused regardless of obesity. Type 1 diabetes can occur in childhood age.

TYPE 2- In this case the various organs of the body become insulin resistant, and this increases the demand for insulin or fails to produce insulin. Type 2 generally occurs in the middle age groups.

GESTATIONAL DIABETES- It is a type of diabetes that tends to occur in pregnant women due to the high sugar levels as the pancreas don't produce sufficient amount of insulin.

- General symptoms of diabetes,

1. Increased thirst
2. Increased urination - Weight loss

3. Increased appetite - Fatigue
4. Nausea and/or vomiting - Blurred vision
5. Slow-healing infections - Impotence in men

- Diagnose test,

1. Urine test
2. Fasting blood glucose level
3. Random blood glucose level
4. Oral glucose tolerance test
5. Glycosylated hemoglobin

Diabetes patients can often loss of sensation in their feet. Even the smallest injury can cause infection that can be very dangerous. 15% of the patients with diabetes will develop foot ulcers due to nerve damage and reduced blood flow. Diabetes slowly steals the person's vision. It is the cause for common blindness and cataracts.

II.RELATED WORK

In data mining various algorithm and technique used for study and analysis of various disease like cancer, hepatitis, thyroid etc.in recent survey show that diabetes is one of the most fast growing disease.
Orabi et al. in [5] designed a system for diabetes prediction, whose main aim is the prediction of diabetes a candidate is suffering at a particular age. The proposed system is designed based on the concept of machine learning, by applying decision tree. Obtained results were satisfactory as the designed system works well in predicting the diabetes incidents at a particular age, with higher accuracy using Decision tree.
Pradhan et al in [2] used Genetic programming (GP) for the training and testing of the database for prediction of diabetes by employing Diabetes data set which is sourced from UCI repository. Results achieved using Genetic Programming gives optimal accuracy as compared to other implemented techniques. There can be significant improve in accuracy by taking less time for classifier generation. It proves to be useful for diabetes prediction at low cost. There are many ways to predict multilevel disease. The main aim of this paper is to find out best classifier from different classification algorithm that can be used to predict disease on applying data set of the patients [3].
The research paper, "A 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 [4]. 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 [1]. This paper also tells about predictive and descriptive type about the data. This paper describes about Diseases Prediction; Classification algorithm; Data Mining, Decision tree. The main aim of this paper is to find out best classifier from different classification algorithm that can be used to predict disease on applying data set of the patients.

III. DATA MINING TECHIQUES

A.DATA MINING ALGORITHMS

1 .Naive Bayes classifier:

Naïve Bayes is a classification technique with a notion which defines all feature are independent and unrelated to each other. Naïve Bayes can be a powerful predictor. This technique very useful for large datasets. Naive Bayes is a machine learning classifier which employs the Bayes theorem. Naive Bayes is known to beat even profoundly advanced grouping techniques. Bayes theorem provides a simple method of calculating posterior probability P (c in x) from P(c), P(x) and P (x in c).

Look at the equations below:

P (C in X) = P (X in C) P(C)/P(X)
P (C in x) is the posterior probability of class (C, target) given predictor (x, attributes).
P (C) is the prior probability of class.
P (X in C) is the likelihood which is the probability of predictor given class.
P (X) is the prior probability of predictor. The evaluated performance of Naive Bayes algorithm using Confusion Matrix is as follows:

Table 1.Confusion Matrix of Naïve Bayes

	A	B
Tested negative	422	78
Tested positive	104	168

2. Support Vector Machine (SVM):

SVM is one of the standard set of supervised machine learning model employed in classification. Given a two-class training sample the aim of a support vector machine is to find the best highest-margin separating hyperplane between the two classes [7]. For better generalization hyperplane should not lies closer to the data points belong to the other class. Hyperplane should be selected which is far from the data points from each category. The points that lie nearest to the margin of the classifier are the support vectors. The Accuracy of the experiment is evaluated using WEKA interface. The SVM finds the optimal separating hyperplane by maximizing the distance between the two decision boundaries. Mathematically, we will maximize the distance between the hyperplane which is defined by $wT\ x + b = -1$ and the hyperplane defined by $wT\ x + b = 1$ this distance is equal to 2 _*w*_. This means we want to solve max 2 _*w*_. Equivalently we want min _*w*_| 2. The SVM should also correctly classify all x (i), which means *yi* (*wT xi* + *b*) >= 1, _*i* _ {1, ¢¢, *N*}. The evaluated performance of SVM algorithm for prediction of Diabetes, using Confusion Matrix is as follows:

Table 2. Confusion Matrix of SVM

	A	B
Tested negative	500	0
Tested positive	268	0

3. Decision Tree Classifier:

Decision Tree is a supervised machine learning algorithm used to solve classification problems. The main objective of using Decision Tree in this research work is the prediction of target class using decision rule taken from prior data. It uses nodes and internodes for the prediction and classification. Root nodes classify the instances with different features. Root nodes can have two or more branches while the leaf nodes represent classification. In every stage, Decision tree chooses each node by evaluating the highest information gain among all the attributes [6]. The evaluated performance of Decision Tree technique using Confusion Matrix is as follows:

Table 3. Confusion Matrix of Decision Tree

	A	B
Tested negative	407	93
Tested positive	108	160

IV. DATASET USED

In this system we used WEKA tool for performing the experiment WEKA is a software which is designed in the country New Zealand by University of Waikato, WEKA is free software available under the GNU General Public License. The Weka is a collection of machine learning algorithms for solving real-world data mining problems. The WEKA is a workbench contains a collection of visualization tools and algorithms for data analysis and predictive modeling, together with graphical user interfaces for easy access to this functionality which includes a collection of various machine learning methods for data classification, clustering, regression, visualization etc. The main aim of this study is the prediction of the patient affected by diabetes using the WEKA tool by using the medical database PIDD. Table-4 shows a brief description of the dataset.

Table 4. Dataset Description

Database	No. of attribute	No. of instances
PIDD	8	768

PIDD –PIMA INDIAN DIABETES DATASET:

The proposed methodology is evaluated on Diabetes Dataset namely (PIDD), which is taken from UCI Repository. This dataset comprises of medical detail of 768 instances which are female patients. The dataset also comprises numeric-valued 8 attributes where value of one class '0' treated as tested negative for diabetes and value of another class '1' is treated as tested positive for diabetes. Dataset description is defined by Table-4 and the Table-5 represents Attributes descriptions. The World Health Organization proposed these attributes, depicted below in Table, of physiological measurements and medical test results for the diabetes diagnosis. PIDD database plays an important role in prediction of multilevel disease.

Table 5. Attribute Description

Sl. no.	Attribute	Abbreviation
1	Number of times pregnant	Pr
2	Plasma glucose concentration	Pl
3	Diastolic blood pressure	Pr
4	Triceps skin fold thickness (mm)	sk
5	2-hr serum insulin (μU/ml)	in
6	Body mass index	ma
7	Diabetes pedigree function	pe
8	Age in years	ag
9	Class'0'or'1'(0-healthy, 1-diabetes)	cl

V. SIDE EFFECT OF DIABETES

In addition to the symptoms, diabetes can cause long term damage to our body diabetes affects our blood vessels and nervous and heart attack and stroke therefore can affect any part of the body. The risk is greater for people with diabetes, who have progressed cholesterol and blood pressure levels. If the family has diabetic history also increases heart problems. To reduce the risk and pick up any problems early: Have the blood pressure checked at least every six months, but more often if person have high blood pressure or are taking medication to lower this. Have the test of HbA1c checked at least every year it may need to be checked three to six monthly. Have the cholesterol checked at least yearly. Further pathology tests such as an electrocardiogram (ECG) or exercise stress test may also be recommended by doctor. Heart disease and blood vessel are common problems for many people who don't have their diabetes under control. Blood vessel damage and nerve damage may also cause foot problems that, in rare cases, can lead to amputations. People having diabetes are ten times occurred cause to damage the whole body.

VII .CONCLUSION

In this paper the various data mining techniques are used to predict the person is diabetes or not. Using data mining techniques the healthcare management predicts the disease and diagnosis of diabetes. In this we used Naïve Bayes, SVM classifier and decision tree classifier for better prediction of diabetes disease. In future we used more attributes for prediction. In the data mining methods to aid people to predict diabetes has gain major popularity. The Naïve Bayes has highest accuracy to predict the person is diabetic or not as compare to SVM and Decision tree classifier. All above methods used to predict diabetes. But if the Patient is detected as diabetes firstly there is a need of finding Control and Un-control condition of diabetes. Because if Patient has diabetes in Un- control condition, may be the patient has severe effect on Patient's Organ like Heart, Eye, Kidney etc. So there is need of finding early Stage which may be help patient for reducing the Severity on Organ or Halting the Severe Effect on Organ.

VIII.REFERENCES

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

[2] Bamnote, M.P., G.R., 2014. "Design of Classifier for Detection of Diabetes Mellitus Using Genetic Programming Advances in Intelligent Systems and Computing, 763–770.

[3] Isha Vashi, Prof. Shailendra Mishra, "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.

[4] Nilesh Jagdish Vispute, Dinesh Kumar Sahu, Anil Rajput,"A Survey 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.

[5] Orabi, K.M., Kamal, Y.M., Rabah, T.M., 2016.Early Predictive System for Diabetes Mellitus Disease, in: Industrial Conference on Data Mining, Springer.Springer.pp.420–427.

[6] Iyer, A., S, J., Sumbaly, R., 2015. Diagnosis of Diabetes Using Classification Mining Techniques. International Journal of Data Mining & Knowledge Management Process 5, 1–14.

[7] Sisodia, D., Srivastava, S.K., Jain, R.C., 2010.ISVMfor face recognition. Proceedings2010 International Conference on Computational Intelligence and Communication Networks, CICN2010, 554–559.