

Estimation and Prediction of Diabetes Mellitus Using Association Summarization Techniques

Meena

Assistant Professor, Dept. of CSE, PEC, Chennai, India

B.Sindhupriya

B.E. Student, Dept. of CSE, PEC, Chennai, India

G.Sowmya

B.E. Student, Dept. of CSE, PEC, Chennai, India

ABSTRACT

Diabetes Mellitus is a chronic disease, in extreme cases it may leads to death. The growth of Diabetic patients are increasing day by day due to various causes such as obesity, bad diet, pollution etc. In Health Care Systems, the diagnosis of Diabetes has been comprehensively investigated showing acceptable levels of accuracy. Our motive is to design a model which can prognosticate likelihood of Diabetes in patient's with maximum accuracy. The data set taken from PIDD are used as train data and data collected from hospitals are taken as test data. The objective of the project is to predict the diabetes using data mining algorithms and to compare its accuracy. Performance comparison of Decision tree, Naïve Bayes, SVM algorithms has been done on Pima Indian Diabetes Dataset. The highest accuracy measure was calculated using instances.

Keywords - Diabetes, accuracy, Naïve Bayes, Decision Tree, SVM.

I. INTRODUCTION

Diabetes is a heterogeneous group of diseases characterized by chronic elevation of glucose in the blood. It arises because the body is unable to produce enough insulin of its own needs, either because of impaired insulin secretion, impaired insulin action or both. The three classic symptoms of Diabetes are thirst and weight loss. The number of people with Diabetes has risen from 108 million in 1980 to 422 million in 2018. Data sets involving Artificial Intelligence, Machine Learning, Statistics and Database Systems. Its main aim being to extract information from a data set and transform it into a understandable structure for further usage.

II. MOST COMMON TYPES

2.1 Type 1 Diabetes

A chronic condition in which the pancreas produces little or no insulin. It usually starts in childhood or young adulthood. It happens when your immune system destroys cells in your pancreas called beta cells. This condition cannot be cured.

2.2 Type 2 Diabetes

A chronic condition that affects the way the body produces blood sugar(glucose). The body either does not produce enough insulin, or it resists insulin. It can develop at any age and usually discovered during adulthood.

2.3 Type 3 Diabetes

A form of high blood sugar affecting pregnant women. A blood sugar test during pregnancy is used for diagnosis. Those who develop gestational diabetes are at the high risk of developing type 2 diabetes. Treatable by a medical professional.

III. OVERVIEW

3.1 Dataset

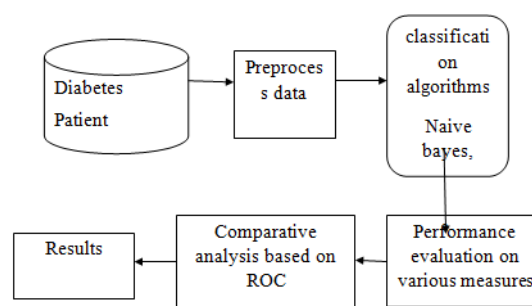
The proposed system is processed on PIDD which is taken from UCI repository. The dataset contains 8 attributes which includes blood pressure, skin thickness, BMI, diabetic pedigree function, age, insulin, number of pregnancies, glucose.

0= tested negative

1= tested positive

IV. PROPOSED SYSTEM

In this system we focus in Electronic Health Records(EHR) to discover set of risk factors that represents a patients with the risk of developing diabetes. In our proposed model we compare the accuracy of prediction algorithms in data mining. Accuracy can be calculated using Precision, F-measure and Recall values. Naïve Bayes classifier reduces the misclassification rate.



V. RELATED WORKS

Prediction of Diabetes mellitus using classification algorithms

- The objective of DeeptiSisodia and Dilip Singh Sisodia was to predict the Diabetes with maximum accuracy. The different data mining algorithms like SVM, Naïve Bayes, Decision tree are used to predict the disease at an early stage.
- Prediction of Diabetes using data mining techniques.
- The objective of Dr. D. I. George Amalarethnam and N. Aswignesh was to compare the different techniques of classification algorithms such as Naïve Bayes, Decision tree, K-means, Support vector Machines, Apriori.
- Prediction of Diabetes using data mining techniques.
- The objective of v. Mareeswaari, Saranya R, Mahalakshmi R and Preeti E was to diagnose whether the person is affected by Diabetes or not using k-nearest neighbor classification algorithm.
- Study of Data Mining algorithms for the prediction and Diagnosis of Diabetes Mellitus.
- The objective of Veena Vijayan V, Aswathy Ravikumar is to diagnose the disease using EM algorithm, KNN algorithm, ANFIS algorithm, K-means algorithm. Here the KNN algorithm is used for classifying the objects.
- Various Data Mining techniques for Diabetes Prognosis.
- The objective of Misha Reyaz, Gagan Dhawan is to diagnose Diabetes Mellitus at an early stage using various data mining approaches.

VI. MODULE DESCRIPTION

6.1 Preprocessing

The collected data must be preprocessed before applying data mining concepts. The datasets are pre-processed to analyze the class labels. Data cleaning eliminates the noisy data and replaces missing data from target dataset, it also removes the duplicate.

6.2 Training Phase

The classification algorithm is now applied to cleaned dataset. The Naïve Bayes and the Decision tree Classification algorithms are now applied and it classifies correctly classified and incorrectly classified instances.

6.3 Testing Phase

In this phase by using Naïve Bayes algorithm the misclassification rate is reduced. The Dataset is applied on this algorithm and tested and therefore predicted that the patient is suffering from diabetes or not.

6.4 Performance Measure

ACCURACY

Accuracy=(TN+TP)/total instances

PRECISION

Precision=TP/(TP+FP)

RECALL

zRecall=TP/(TP+FN)

F-MEASURE

F-Measure=2PR/(P+R)

VII. CONCLUSION

Overall the predicted system has achieved high accuracy and the misclassification rate got reduced. It is evident that the middle age people and the aged people are mostly affected by diabetes. Hence it is noticed that as the age progresses the probability of getting diabetes increases.

REFERENCES

- [1] Prediction of Diabetes using classification Algorithms.(2018) Deepti Sisodia, Dilip Singh Sisodia, National Institute of Technology, Rajpur, India.
- [2] Prediction of Diabetes using Data Mining Techniques.(2017) V.Mareeshwari, Saranya R, Mahalakshmi R, Preethi E, Department of Software Engineering, VIT University.
- [3] Study of Data Mining algorithms for Prediction and Diagnosis of Diabetes Mellitus.(2016) Veena Vijayan V, Aswathy Ravikumar, Department of Computer Science, Mar Baselios College of Engineering and Technology.
- [4] Prediction of Diabetes Mellitus using Data Mining techniques.(2016) Dr.D.I.George Amalarethnam, N. Aswin Vignesh, Department of Computer Science and Engineering, Government College of Arts and Science.
- [5] Various Data Mining techniques for Diabetes Prognosis.(2017) Misha Reyaz, Gagan Dhawan, Modern Institute of Engineering and Technology, Kurukshetra University.