



iJRASET

International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 7 Issue: III Month of publication: March 2019

DOI: <http://doi.org/10.22214/ijraset.2019.3110>

www.ijraset.com

Call: ☎ 08813907089

E-mail ID: ijraset@gmail.com

Estimation and Prediction of Diabetes Mellitus using Association Summarization Technique

J. Omana¹, D. C. Pravalika², R. Sowmiya³

¹Assistant Professor, ^{2,3}Student, Prathyusha Engineering College, Thiruvallur

Abstract: Diabetes mellitus is the one of the world's major diseases which causes increases in blood sugar level. The risk of diabetes is increasing day by day, and many complications may occur if diabetes remains untreated and unidentified. The diagnosis of diabetes is tedious process. But the rise in machine learning approaches solves this critical problem. Machine learning classification algorithms are used in this model to detect diabetes at early stage. The performance of algorithms are measured using Precision, Accuracy, F-measure, and Recall. These results are verified using ROC curve in a proper and systematic manner.

I. INTRODUCTION

Diabetes is metabolic disease which affects the ability of the body in producing hormone insulin, which in turn raise the levels of glucose in the blood. In diabetes a person generally suffers from high blood sugar. Intensify thirst, Intensify hunger and frequent urination are some of the symptoms caused due to high blood sugar. The disease is of three types: insulin-independent, insulin-dependent and gestational diabetes. The insulin-dependent is known as Type 1 DM which is due to lack of insulin production. This is mostly seen in children. The insulin-independent also known as Type 2 DM, is due to the failure of cells to respond to insulin produced. People who have obesity will have type 2 diabetes. The third type which gestational diabetes occurs in women during gestation period. Blood pressure, plasma glucose may the reason for diabetes.

Many researchers are conducting experiments for diagnosing the disease using various classification algorithms of machine learning approaches and then machine learning algorithms works better in diagnosing different diseases. Data mining and machine learning algorithms gain its strength due to capability of managing large amount of data from several different sources and integrating the background information in the study. Machine learning algorithms are used and evaluated on PIDD dataset to find prediction of diabetes in a patient.

II. LITERATURE SURVEY

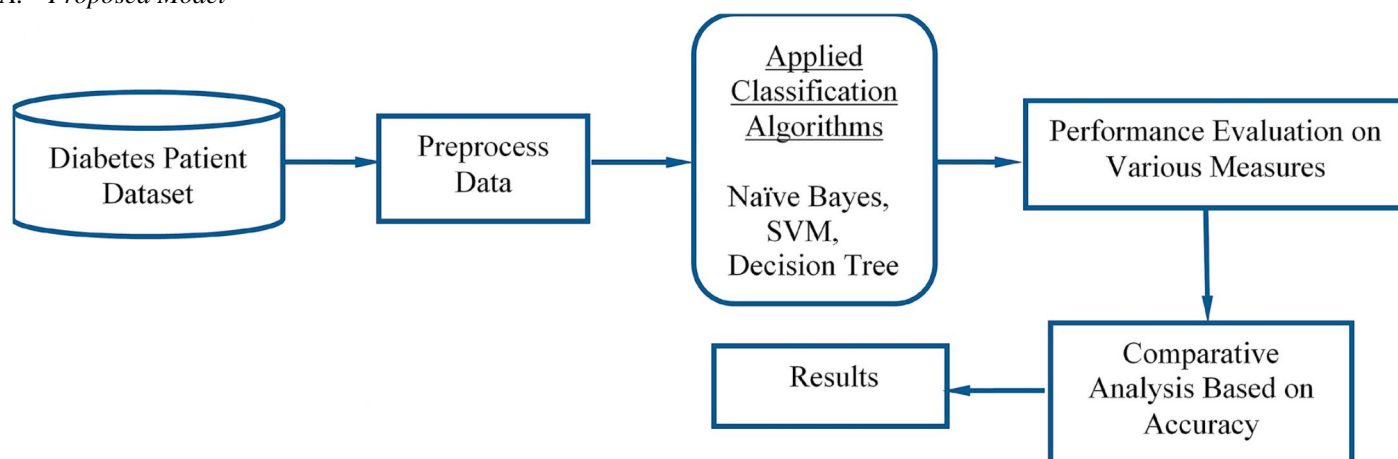
- 1) *KNN Algorithm:* KNN is a simple and also a lazy learning algorithm. It is one of the classification algorithm used in health care. It can be used for both classification and regression. It is more widely used in classification problems. The algorithm is preferred mostly for its ease of interpretation. KNN is distance weighted and majority voting algorithm.
- 2) *Naïve Bayes Classifier:* It is a classification technique with a notion which defines all features are independent and unrelated to each other. It defines the status of a specific feature in a class does not affect the status of another feature. Since it is based on conditional probability it is considered as powerful algorithm employed for classification purpose. It works well for data with problems an missing values. Naïve bayes is a machine learning classifier which employs Bayes Theorem.
- 3) *Decision Tree Classifier:* It is a supervised machine learning algorithm used to solve classification problems. It can predict target class using decision rules taken from prior data. It uses nodes and internodes for prediction and classification. Root nodes classify instances with different features. Root nodes can have two or more branches while the leaf nodes represent classification.

III. IMPLEMENTATION

Steps To Compute Process In Prediction Of Diabetes Mellitus

- 1) *Preprocessing:* As the first step the target data must be collected before applying data mining concepts. The datasets are pre-processed to analyze the class labels. Data cleaning removes noisy data and replaces missing data from target dataset.
- 2) *Training Phase:* The classification algorithm is now applied to cleaned dataset. The algorithm now classifies correctly classified and incorrectly classified instances. The classifier accuracy is tested for different values of parameter k.
- 3) *Testing Phase:* The distance between target data which is unknown and each instance of diabetes dataset which is known is found using Euclidean distance measures. The computed distances are sorted and closest of target class are considered as per the parameter k. The majority of these class labels are assigned to target variable. This predicts whether the person has diabetes or not.

A. Proposed Model



IV. ALGORITHMS USED

A. 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[26]. 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 .

B. Naive Bayes Classifier

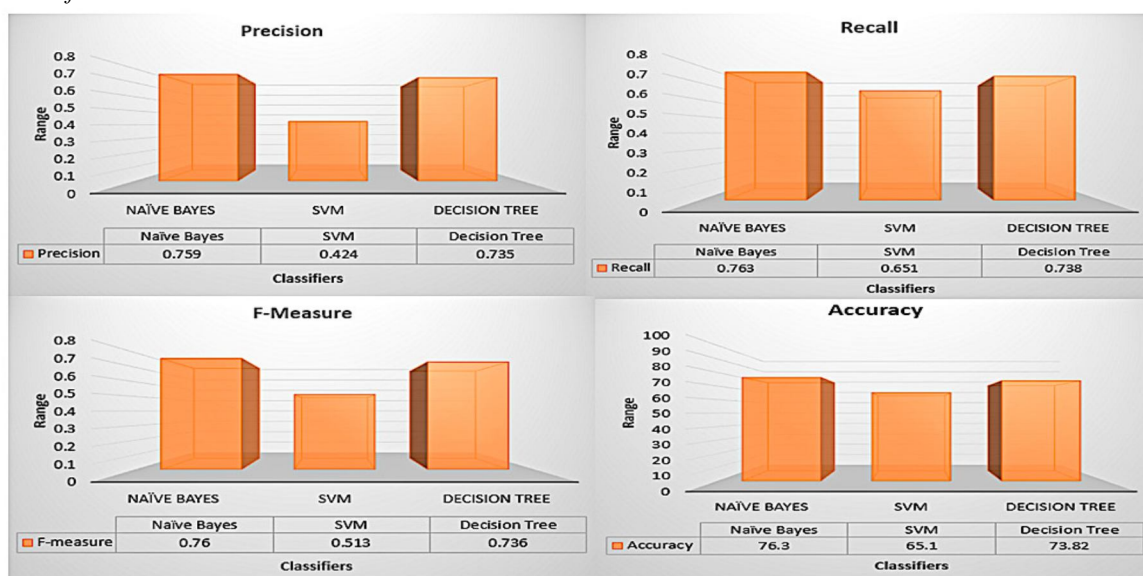
Naive Bayes is a classification technique with a notion which defines all features are independent and unrelatedto each other. It defines that status of a specific feature in a class does not affect the status of another feature. Sinceit is based on conditional probability it is considered as a powerful algorithm employed for classification purpose.

C. Decision Tree Classifier

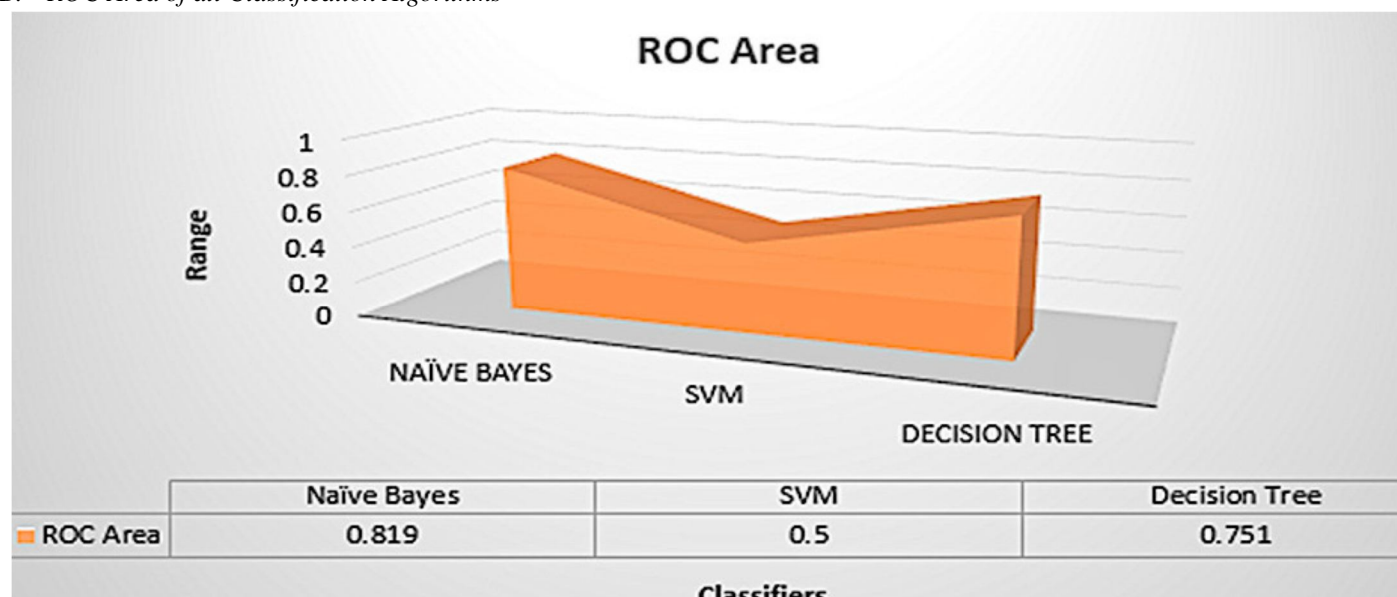
Decision Tree is a supervised machine learning algorithm used to solve classification problems. The main objective of using Decision Tree in this research work is the prediction of target class using decision rule taken from prior data. It uses nodes and internodes for the prediction and classification.

V. RESULTS

A. Classifier Performance on Various Measures



B. ROC Area of all Classification Algorithms



VI. CONCLUSION

One of the important real-world medical problems is the detection of diabetes at its early stage. In this study, systematic efforts are made in designing a system which results in the prediction of disease like diabetes. During this work, three machine learning classification algorithms are studied and evaluated on various measures. Experiments are performed on Pima Indians Diabetes Database. Experimental results determine the adequacy of the designed system with an achieved accuracy of 76.30 % using the Naive Bayes classification algorithm. In future, the designed system with the used machine learning classification algorithms can be used to predict or diagnose other diseases. The work can be extended and improved for the automation of diabetes analysis including some other machine learning algorithms.

REFERENCES

- [1] Prediction of Diabetes using Data Mining Techniques.(2017) V.Mareeshwari, Saranya R, Mahalakshmi R, Preethi E, Department of Software Engineering, VIT University.
- [2] 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.
- [3] 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.
- [4] Various Data Mining techniques for Diabetes Prognosis.(2017) Misha Reyaz, Gagan Dhawan, Modern Institute of Engineering and Technology, Kurukshetra University.



10.22214/IJRASET



45.98



IMPACT FACTOR:
7.129



IMPACT FACTOR:
7.429



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24*7 Support on Whatsapp)