TEST Engineering & Management

January-February 2020
ISSN: 0193-4120 Page No. 6597 - 6600

Predictive Measure of Diabetes for Early Detection using Data Mining Approaches

A. S Pawan Kalyan[1], R. Senthil Kumar[2]

[1]UG Student, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences
[2]Assistant professor, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences
[1]pawankalyannirmala@gmail.com, [2]rsenthilmecse@gmail.com

Article Info
Volume 82
Page Number: 6597 - 6600
Publication Issue:
January-February 2020

Abstract
Diabetes is one of the major non-communicable diseases which has a great impact on human day to day activities. Diabetes is an infection that happens when your blood glucose, additionally known as glucose, is excessively high. Blood glucose is your primary wellspring of vitality and originates from the nourishment you consume. Insulin, a hormone made by the pancreas, enables glucose from nourishment to get into your cells to be utilized for vitality. The body makes the insulin each and every time, but it will not utilize the insulin fully. Glucose at that point remains in your blood and doesn't arrive at your cells. After some time, having an excessive amount of glucose in your blood can mess wellbeing up. Despite the fact that diabetes has no fix, you can find a way to deal with your diabetes and remain sound. Data mining approach analyze patient's ailments. Diabetes Mellitus is an incessant sickness to influence different organs of the human body. Early forecast can save human life and can assume responsibility for the maladies. Choosing genuine classifiers plainly grows the accuracy and skill of the framework. Because of its persistently expanding rate, an ever increasing number of families are out of line by diabetes mellitus. Most diabetics think minimal about their hazard factor they face preceding analysis. This paper investigates the early forecast of diabetes utilizing data mining methods.

Keywords: *glucose, diabetes, insulin, data mining*

Article History
Article Received*: 18 May 2019*
Revised*: 14 July 2019*
Accepted*: 22 December 2019*
Publication*: 01 February 2020*

1. Introduction

According to world health association which used to collect the details of the persons who have experiencing diabetes and they will add the records up to 420 million per year. Each year there is an increase in number of persons who have diabetes and the details are recorded in many clinical centres. The world wellbeing association (WHO) covers "Diabetes Care 2018" by American Diabetes Association and Standards for Medical consideration in Diabetes, an examination for connection different races and their compensation. The various people (sexual orientation and compensation) developed somewhere in the range of 29 and 70 years, level of passing due to hypertension. Diabetes is a type of chronic Disease which is caused by the high sugar level in the blood circulatory system. It is even caused by improper function of pancreatic beta cells and it will affects the other parts of the body like kidney failure, risk of heart failure, foot issues , eye visual blindness and nerve issues. According to the Fig1 diabetes is most commonly seen in the people who belong to low income category particularly in males.

The proposed solution for this type problem can be solved using data mining and classification techniques are prediction algorithms. These algorithms has the capability to predict the disease in a right way

January-February 2020
ISSN: 0193-4120 Page No. 6597 - 6600

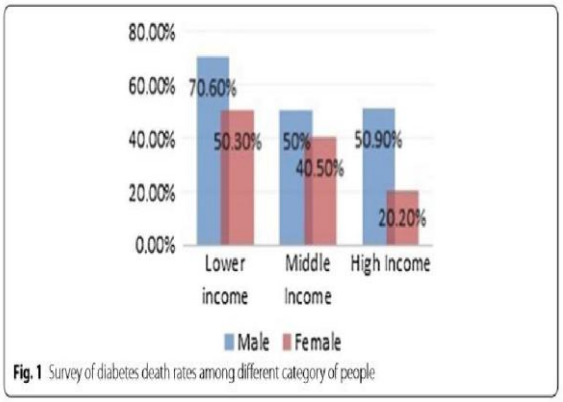


Figure 1: Survey of diabetes death rates among different category of people

The algorithm which is used for the predictions are:

- Support vector machine
- K nearest Neighbour
- Random forest

Support vector machine

Support vector machines is a process characterized as directed learning models that are related with learning calculation analysis for classification and regression analysis. The SVM algorithm is used builds the training model and arrange it in a category or other probabilistic binary classifier.

K nearest neighbour

The algorithm of (KNN) is a kind of supervised machine learning which is used in both classification and regression model.

The main function of this algorithm is to take the labelled data and produce the output based on the learning of unlabelled data.

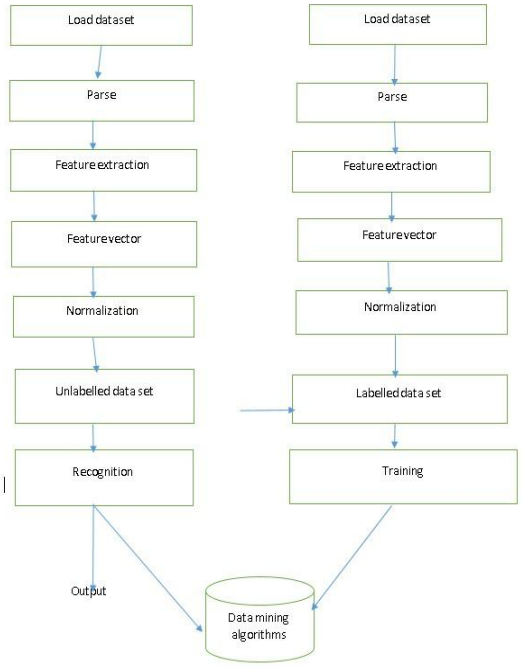
Random Forest

It is an algorithm which uses ensemble learning methods for construction of decision trees. This decision trees are used for classification and regression problems where it takes the input from training data and produce the output class based on the mode of classes or for the individual trees.

2. Literature Review

Based on the sources from international diabetes federation that every one out of eleven adults have diabetes in 2015 all over the globe. According to the sources that in next 25 years there are chances that getting diabetes for men and women will increase more than the current analysis. Using the data mining techniques where we can employ the prediction algorithms, so it will help the persons to take preventive actions from the diagnosis results from these methods suggest that the large dataset will be more efficiency to the overall result of the system. But there is an advantages in small data set, because it will work more efficiently and effectively on the accuracy of the system it will be useful to number of medical tests because it is a less time consuming process compared to the previous method for the conducting the medical tests. The algorithm used in their system is (KNN) K nearest neighbour which is a lazy learning and a non-parametric algorithm where it will use several points of classes to create a classification of new sample point. But the main drawback in this algorithm is that it will not learn the training data, so it is not suitable for using it for prediction of diabetes disease.

This analysis would suggest that the methods used in the diabetes prediction using matlab tools. The matlab tools is one of the most used application in artificial intelligence, but these system are designed to gives results based on the part inferences. The algorithm that is used here in the place of pin blood glucose level devices and it is replaced since it is a painful procedure, but the result is accurate only in this type of devices. The algorithm that is used here is (BPNN) back propagation neural network which is used to the derivatives to calculate in less time and it is an algorithm comes under artificial intelligence. The algorithm which will learn and compute the results or weight with respect to gradient descent where it will travel from bottom to the top. But the main disadvantages of this algorithm is slow and inefficient which can even get stuck in the local minima during the time of prediction process.



3. Existing System

Few years back, all the details are recorded manually based on the various tests and techniques that are used for the treatment of diabetes. This techniques and materials that are used have some issues in India. Many people have created different classification model to early prediction of diabetes using the data mining techniques.

January-February 2020
ISSN: 0193-4120 Page No. 6597 - 6600

Grouping and classifying the hereditary neural system, which is used in diabetes informational collections is used to managing the missing and anomaly data. The missing data are supplanted using the space with the comparing property. The neural system is capable to forecast the pre-prepared dataset, In present techniques of diabetes treatment which using the data mining strategies to the design will perform computational calculations that will be helpful for the treatment of diabetes From these experiments we observe that among youthful persons have less chance of getting diabetes whereas the senior or old persons require treatment immediately. This analysis is been effective among the diabetic informational collection and provides greater intention to growing expectation models. Data mining will be effective for this type of predicting calculations.

4. Proposed System

Since the data that is growing in healthcare organizations in an unstructured way, we need some system to manage the data in a structured way with a possible expected solution. In this paper the data is disintegrated using R tool and the numerical are converted in nominal values and the datasets are divided in two types and they are training and testing the datasets. These datasets are used to create a data model. These datasets can be loaded in the prediction algorithms and generates the data models. These data models are used in test set and the accuracy is measured for each data model. This attempt is made to classify the diabetic dataset that is loaded on the console and the accuracy of various models compared.

5. Results

The study was carried out through diabetic dataset to evaluate the awareness of diabetics. The finding of our study may help physician to focus on helping the family members of diabetic patients on personal lifestyle modifications that will prevent the diabetes or postpone the disease complications in the populations. The overall objective of this study was to predict and analyse the performance (figure 2) of various prediction algorithm and knowledge of various aspects of diabetes and its complications. More than 91% of proposed work will find diabetics in efficient way compared to existing system**.**

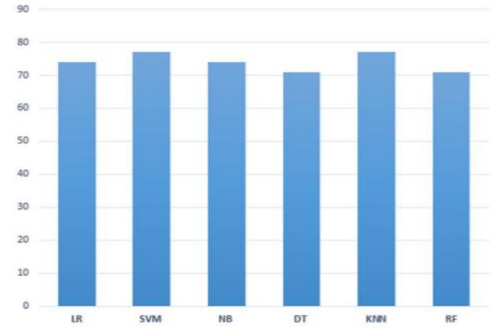


Figure 2: Accuracy of algorithms

6. Conclusion

Using Data Analytics in data mining usage gives orderly approach to accomplishing better results like accessibility and moderateness of medicinal services administration to all populace. NonCommunicable Diseases like diabetes, is one of a noteworthy wellbeing peril in India. By changing different wellbeing records of diabetic patients to helpful dissected outcome, this examination will influence the patient to comprehend the intricacies to happen. The objective of this exploration manages the investigation of diabetic treatment in human services industry utilizing enormous information examination. The plan of prescient investigation arrangement of diabetic treatment may give improved information and examination yield the best outcomes in social insurance. By utilizing area mindful human services administration, anybody from rustic territory can get appropriate treatment with ease. This examination primarily engaged for the patients in the country zone. Treatment can be offered when it is recognized ahead of time. The Best performing algorithm will be used for prediction of diabetes.

References

[1] A. Belle, R. Thiagarajan, S. M. R. Soroushmehr, F. Navidi, D. A. Beard, and K. Najarian, “Big Data Analytics in Healthcare,” Hindawi Publ. Corp., vol. 2015, pp. 1–16, 2015.

[2] J. Andreu-Perez, C. C. Y. Poon, R. D. Merrifield, S. T. C. Wong, and G.-Z.Yang, “Big Data for Health,” IEEE J. Biomed. Heal. Informatics, vol. 19, no. 4, pp. 1193–1208, 2015

[3] E. Ahmed et al., “The role of big data analytics in Internet of Things,” Comput. Networks, vol. 129, no. December, pp. 459–471, 2017

[4] M Panda and M Patra, “Network Intrusion Detection Using Naïve Bayes” at IJCSNS International Journal of Computer Science and Network Security, VOL7, 2007.

[5] A.Sangasoong song and J.Chongwatpol, “An analysis of diabetes risk factors using data mining approach,” Oklahoma State University, Stillwater, OK 74078, USA.

[6] SadeghB. Imandoust, M.Bolandraftar, “Application of Nearest Neighbor (KNN) Approach for Predicting Economic Events: Theoretical Background”, International Journal of Engineering Research and Applications, Vol. 3, 2013.

[7] L.Wang, L.Khan and B.Thuraisingham, “An Effective Evidence Theory based on nearest Neighbor (KNN) classification”, IEEE International Conference, 2008.

[8] H. Schneiderman and T. Kanade. A statistical method for 3d detection applied to faces and cars. In Proceedings of CVPR2000, 2000.

January-February 2020
ISSN: 0193-4120 Page No. 6597 - 6600

[9] K Beyer, J Goldstein, R Ramakrishna and U Shaft," When is 'Nearest neighbor' Meaningful?" 2014.

[10] Chitra, A. & Uma, S. (2010) "An Ensemble Model of Multiple Classifiers for Time Series Prediction", International Journal of Computer Theory and Engineering, 2(3): 1793-8201.

[11] Milan Kumari, SunilaGodara, "Comparative Study of Data Mining Classification Methods in cardiovascular Disease Prediction", IJCST, Vol. 2, Issue 2, 2011, pp. 304-308.

[12] W. H. S.Gunarathne, K. D. .Perera, and K. A. D. C.Kahandawaarachchi, "Performance Evaluation on Machine Learning Classification Techniques for Disease Classification and Forecasting through Data Analytics for Chronic Kidney Disease (CKD)," in 2017 IEEE 17th International Conference on Bioinformatics and Bioengineering (BIBE), 2017, pp. 291–296.