



IJIRCCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 6, June 2022

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.165



9940 572 462



6381 907 438



ijircce@gmail.com



www.ijircce.com

Diabetes Prediction Using Machine Learning

¹D.HARSHITHA ².M.HARSHITHA CHOWDARY, ³.G.SIREESHA, ⁴.J.SREE LASYA

^{1,2,3,4} B. Tech Students, Department of Information Technology, Vasireddy Venkatadri Institute of Technology, Guntur, Andhra Pradesh, India

ABSTRACT: Diabetes is an illness caused because of high glucose level in a human body. Diabetes should not be ignored if it is untreated then Diabetes may cause some major issues in a person like: heart related problems, kidney problem, blood pressure, eye damage and it can also affect other organs of human body. Diabetes can be controlled if it is predicted earlier. To achieve this goal, this project work will do early prediction of Diabetes in a human body or a patient for a higher accuracy through applying various Machine Learning Techniques. Machine learning techniques provide better result for prediction by constructing models from datasets collected from patients. This project will use Machine Learning Classification and ensemble techniques on a dataset to predict diabetes. are K-Nearest Neighbour (KNN), Logistic Regression (LR), Decision Tree (DT), Support Vector Machine (SVM) and Random Forest (RF). The accuracy is different for every model when compared to other models. The Project work gives the accurate or higher accuracy model shows that the model is capable of predicting diabetes effectively

KEYWORDS: Diabetes, Decision trees, Knn algorithm, Support vector machines, Diseases, Machine learning algorithms

I.INTRODUCTION

Diabetes is noxious diseases in the world. Diabetes caused because of obesity or high blood glucose level, and so forth. It affects the hormone insulin, resulting in abnormal metabolism of carbs and improves level of sugar in the blood. Diabetes occurs when body does not make enough insulin. According to (WHO) World Health Organization about 422 million people suffering from diabetes particularly from low or idle income countries. And this could be increased to 490 billion up to the year of 2030. However, prevalence of diabetes is found among various Countries like Canada, China, and India etc. Population of India is now more than 100 million so the actual number of diabetics in India is 40 million. Diabetes is major cause of death in the world. Early prediction of disease like diabetes can be controlled and save the human life. To accomplish this, this work explores prediction of diabetes by taking various attributes related to diabetes disease. For this purpose, we use the Pima Indian Diabetes Dataset, we apply various Machine Learning classification and ensemble Techniques to predict diabetes.

Machine Learning is a method that is used to train computers or machines explicitly. Various Machine Learning Techniques provide efficient result to collect Knowledge by building various classification and ensemble models from collected dataset. Such collected data can be useful to predict diabetes. Various techniques of Machine Learning can capable to do prediction, however it is tough to choose best technique. Thus, for this purpose we apply popular classification and ensemble methods on dataset for prediction

II.PROPOSED SYSTEM

Diabetes prediction system is using SVM as proposed system which is a machine learning approach. This algorithm belongs to the family of supervised learning. It is used to train the data. The goal of SVM is to identify an optimal separating hyperplane which maximizes the margin between different classes of the training data. Using SVM we can predict the output whether the person is having diabetes or not

III. PROBLEM STATEMENT

With Machine Learning Techniques, it is possible to predict whether a patient suffers from diabetes or not. We want to build an application to predict whether a person has diabetes or not by considering some attributes like glucose level, insulin, age. This application predicts whether a person is suffering from diabetes.

IV. PROPOSED SYSTEM ARCHITECTURE



Architecture of the system

VI. PROPOSED METHODOLOGY

K-Nearest Neighbour (KNN)

KNN is also a supervised machine learning algorithm. KNN helps to solve both the classification and regression problems. KNN is lazy prediction technique. KNN assumes that similar things are near to each other. Many times, data points which are similar are very near to each other. KNN helps to group new work based on similarity measure. KNN algorithm record all the records and classify them according to their similarity measure. For finding the distance between the points uses tree like structure. To make a prediction for a new data point, the algorithm finds the closest data points in the training data set its nearest neighbors. Here K= Number of nearby neighbors, it is always a positive integer. Neighbors value is chosen from set of class. Closeness is mainly defined in terms of Euclidean distance. The Euclidean distance between two points P and Q i.e. P (p1, p2,..pn) and Q (q1, q2,..qn) is defined by the following equation:- Algorithm- Take a sample dataset of columns and rows named as Pima Indian Diabetes data set.

Take a test dataset of attributes and rows.

Find the Euclidean distance by the help of formula

Then, decide a random value of K. is the no. of nearest neighbors

Then with the help of these minimum distance and Euclidean distance find out the nth column of each. (Find out the same output values. If the values are same, then the patient is diabetic, other- wise not.

Decision Tree-

Decision tree is a basic classification method. It is supervised learning method. Decision tree used when response variable is categorical. Decision tree has tree like structure model which describes classification process based on input feature. Input variables are any types like graph, text, discrete, continuous etc. Steps for Decision Tree

Algorithm-

Construct tree with nodes as input feature.

Select feature to predict the output from input feature whose information gain is highest. (The highest information gain is calculated for each attribute in each node of tree.

Support Vector Machine also known as SVM is a supervised machine learning algorithm. SVM is most popular classification technique. SVM creates a hyperplane that separate two classes. It can create a hyperplane or set of hyperplane in high dimensional space. This hyper plane can be used for classification or regression also. SVM differentiates instances in specific classes and can also classify the entities which are not supported by data. Separation is done by through hyperplane performs the separation to the closest training point of any class.

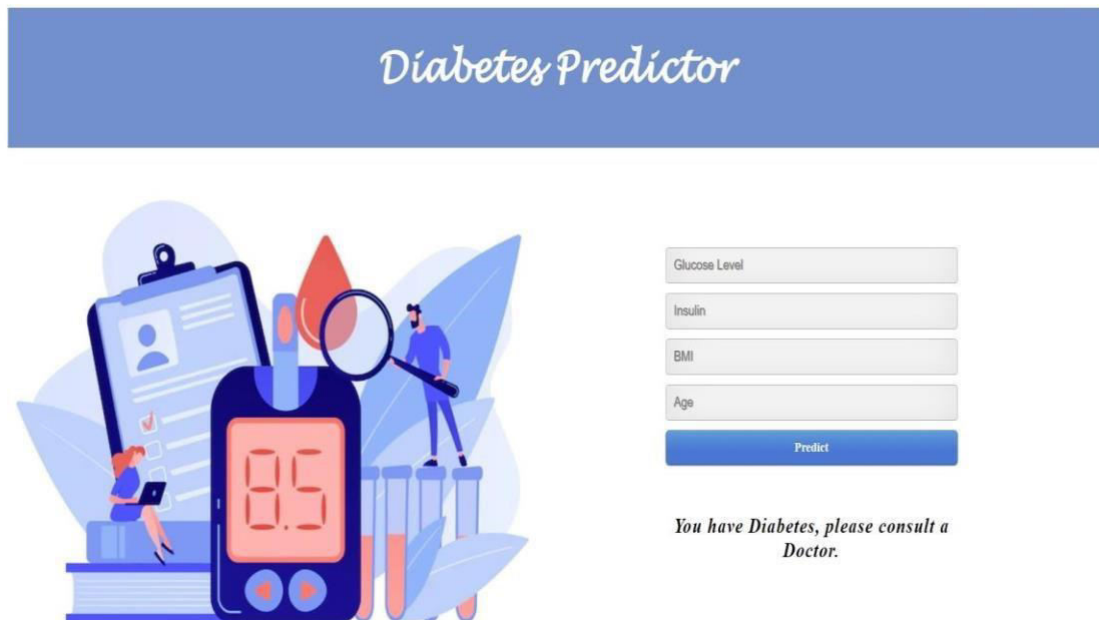
Algorithm-

Select the hyper plane which divides the class better.

To find the better hyper plane you have to calculate the distance between the planes and the data which is called Margin.

VI. RESULTS AND DISCUSSIONS

The following section contains the screenshots of the output and the output is predicted when the user enters the details like age, glucose level, insulin, BMI and then the user can predict whether he/she have the diabetes.



VII. CONCLUSION

So, we used SVM model for the prediction of diabetes which has higher accuracy and recommend the insulin dosage to be taken by a person. For the prediction of diabetes we used SVM model which gives better accuracy compared to the

other machine learning algorithms. A Better model can be developed using neural networks for the prediction of diabetes. Future Research can be done in this area by using Artificial Neural Networks.

REFERENCES

1. B. Nithya and Dr. V. Ilango, "Predictive Analytics in Health Care Using Machine Learning Tools and Techniques", International Conference on Intelligent Computing and Control Systems, 978-1-5386-2745-7, 2017.
2. F. G. Woldemichael and S. Menaria, "Prediction of Diabetes Using Data Mining Techniques," 2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI), 2018, pp. 414-418, doi: 10.1109/ICOEI.2018.8553959.
3. K. Rajesh, V. Sangeetha "Application of Data Mining Methods and Techniques for Diabetes Diagnosis" International Journal of Engineering and Innovative Technology (IJEIT), 2 (3) (September 2012).
4. Debadri Dutta, Debpryo Paul, Parthajeet Ghosh, "Analyzing Feature Importances for Diabetes Prediction using Machine Learning". IEEE, pp 942-928, 2018.
5. Tejas N. Joshi, Prof. Pramila M. Chawan, "Diabetes Prediction Using Machine Learning Techniques". Int. Journal of Engineering Research and Application, Vol. 8, Issue 1, (Part -II) January 2018, pp.-09-13
6. Deeraj Shetty, KishorRit, Sohail Shaikh, Nikita Patil, "Diabetes Disease Prediction Using Data Mining
7. ".International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), 2017 [7] Sisodia, D.; Sisodia, D.S. Prediction of Diabetes using Classification Algorithms. Procedia Comput. Sci. 2018, 132, 1578–1585.
8. Kumari, V.A.; Chitra, R. Classification of diabetes disease using support vector machine. Int. J. Adv. Comput. Sci. Appl. 2013, 3, 1797–180
9. F. A. Khan, K. Zeb, M. AL-Rakhami, A. Derhab, and S. A. C. Bukhari, "Detection and prediction of diabetes using data mining: a comprehensive review," IEEE Access, vol. 9, pp. 43711–43735, 2021. [10] M. K.
10. Hasan, M. A. Alam, D. Das, E. Hossain, and M. Hasan, "Diabetes prediction using ensembling of different machine learning classifiers," IEEE Access, vol. 8, pp. 76516–76531, 2020. [11] American Diabetes Association, "Diagnosis and classification of diabetes mellitus," Diabetes Care, vol. 37, no. 1, pp. 81-90, 2014.



INNO  SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.165

 **doi**[®]
CROSS **ref**

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

 9940 572 462  6381 907 438  ijircce@gmail.com



www.ijircce.com

Scan to save the contact details