International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 05 | May - 2023 SJIF 2023: 8.176 ISSN: 2582-3930

DIABETES PREDICTION USING MACHINE LEARNING

Ms.M.Santhy Computer Science and Engineering & Dhirajlal Gandhi College of Technology
Mr.M.Mohana Priyan Computer Science and Engineering & Dhirajlal Gandhi College of Technology
Mr.M.Syedabuthahir Computer Science and Engineering & Dhirajlal Gandhi College of Technology
Mr.S.Tamilarasan Computer Science and Engineering & Dhirajlal Gandhi College of Technology
Mr.S.Tamilselvan Computer Science and Engineering & Dhirajlal Gandhi College of Technology

---

Abstract - Diabetes is a serious disease that affects the majority of the population. It is caused due to increased blood sugar level because of imbalance in insulin processing by the body. Machine learning is an emerging scientific field in data science dealing with the ways in which machines learn from experience. The aim of this work is to make an early prediction of diabetes more precisely by using a variety of machine learning algorithms. Machine learning algorithms provide better results in diabetes prediction by constructing models from patient datasets. This project has incorporated the algorithms like Naive Bayes, K-Nearest Neighbor (KNN), Logistic Regression, Random Forest, Support Vector Machine (SVM), Decision Tree, Multi-Layer perceptron (MLP) and Logistic Model Tree (LMT). The accuracy is different for all of them.

Key Words: Machine learning, Naive Bayes, K-Nearest Neighbor, Logistic Regression, Random Forest, Support Vector Machine, Decision Tree, Multi-Layer perceptron and Logistic Model Tree.

1.INTRODUCTION

Diabetes is a disease that occurs when your blood glucose, also called blood sugar, is too high. Blood glucose is your main source of energy and comes from the food you eat. Insulin, a hormone made by the pancreas, helps glucose from food get into your cells to be used for energy. Sometimes your body doesn't make enough or any insulin or doesn't use insulin well. Glucose then stays in your blood and doesn't reach your cells.Over time, having too much glucose in your blood can cause health problems. Although diabetes has no cure, you can take steps to manage your diabetes and stay healthy. Sometimes people call diabetes "a touch of sugar" or "borderline diabetes." These terms suggest that someone doesn't really have diabetes or has a less serious case, but every case of diabetes is serious.The most common types of diabetes are type 1, type 2, and gestational diabetes. If you have type 1 diabetes, your body does not make insulin. Your immune system attacks and destroys the cells in your pancreas that make insulin. Type 1 diabetes is usually diagnosed in children and young adults, although it can appear at any age. People with type 1 diabetes need to take insulin every day to stay alive. If you have type 2 diabetes, your body does not make or use insulin well. You can develop type 2 diabetes at any age, even during childhood. However, this type of diabetes occurs most often in middle-aged and older people. Type 2 is the most common type of diabetes. Gestational diabetes develops in some women when they are pregnant. Most of the time, this type of diabetes goes away after the baby is born. However, if you've had gestational diabetes, you have a greater chance of developing type 2 diabetes later in life. Sometimes diabetes diagnosed during pregnancy is actually type 2 diabetes. Less common types include monogenic diabetes, which is an inherited form of diabetes, and cystic fibrosis-related diabetes .

2. SYSTEM IMPLEMENTATION:

EXISTING SYSTEM:

Diabetes Mellitus is among critical diseases and lots of people are suffering from this disease. Age, obesity, lack of exercise, hereditary diabetes, living style, bad diet, high blood pressure, etc. can cause Diabetes Mellitus. People having diabetes have high risk of diseases like heart disease, kidney disease, stroke, eye problem, nerve damage, etc. Current practice in hospital is to collect required information for diabetes diagnosis through various tests and appropriate treatment is provided based on diagnosis. Big Data Analytics plays an significant role in healthcare industries. Healthcare industries have large volume databases. Using big data analytics one can study huge datasets and find hidden information, hidden patterns to discover knowledge from the data and predict outcomes accordingly. In existing method, the classification and prediction accuracy is not so high. In this paper, we have proposed a diabetes prediction model for better classification of diabetes which includes few external factors responsible for diabetes along with regular factors like Glucose, BMI, Age, Insulin, etc. Classification accuracy is boosted with new dataset compared to existing dataset. Further with imposed a pipeline model for diabetes prediction intended towards improving the accuracy of classification.

PROPOSED SYSTEM:

First, the dataset was collected and preprocessed to remove the necessary discrepancies from the dataset, for example, replacing null instances with mean values, dealing with imbalanced class issues, Standardization and Normalization etc. Then the dataset was separated into the training set and test set using the holdout validation technique. Next, different classification algorithms (Logistic Regression, Naive Bayes, Support Vector Machine, K-Nearest Neighbor, Random Forest, Decision Tree, Multi-Layer Perceptron and Logistic Model Tree) were applied to find the best classification algorithm for this dataset. Finally, the best-performed prediction model (Predicting Before 10 Years) is deployed into the proposed website (Using Streamlit).

3.SYSTEM ARCHITECTURE:

The aim of this project is to develop a system which can perform early prediction of diabetes for a patient with a higher accuracy by combining the results of different machine learning techniques. The algorithms like K nearest neighbour, Logistic Regression, Random forest, Support vector machine and Decision tree are used. The accuracy of the model using each of the algorithms is calculated. Then the one with a good accuracy is taken as the model for predicting the diabetes. Fig -1: System Architecture Diagram The presence of disease has been identified using the appearance of various symptoms. However, the methods use different features and produces varying accuracy. The result of prediction differs with the methods/measures/ features being used. Towards diabetic prediction, a Disease Influence Measure (DIM) based diabetic prediction has been presented. The method pre processes the input data set and removes the noisy records. In the second stage, the method estimates disease influence measure (DIM) based on the features of input data point. Based on the DIM value, the method performs diabetic prediction. Different approaches of disease prediction have been considered and their performance in disease prediction has been compared. The analysis result has been presented in detail towards the development.

Fig-1: System architecture

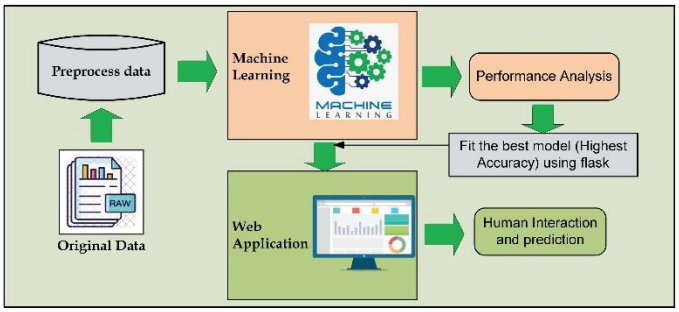

Table -1: Model accuracy after testing

MODEL	ACCURACY
Logistic Regression	78%
K-Nearest Neighbor	87.3%
Naïve Bayes	76.3%
SVM	77.8%
Random Forest	99.6%
Decision Tree	99.2%
MLP	78.7%
Logistic Model Tree	73.6%

4. CONCLUSION:

Diabetes is one of the most chronic and the largest growing disease in India. According to the World Health Organization (WHO), India had 69.2 million people living with diabetes as of 2015. A study conducted by the American Diabetes Association states that India will see a great increase in the number of people diagnosed with diabetes by the 2030. Identifying diabetes or predicting the upcoming of a diabetic life can be propelled by using various machine learning techniques like Naive Bayes, K-Nearest Neighbor (KNN), Logistic Regression, Random Forest, Support Vector Machine (SVM), Decision Tree, Multi-Layer perceptron (MLP) and Logistic Model Tree (LMT) etc. From this project, we can conclude that the best method of prediction of diabetes is Random Forest and Decision Tree. This method gives us an approximate result after the splitting and analysis of the training and testing data. The efficiency of this method is much better compared to that of another all algorithm. The analysis done from the PIMA dataset is really important. The aim of splitting the dataset is to find the highest/best accuracy of the Algorithms and as to how they would respond if the data split set is varied. Procuring the dataset make accuracy of our prediction model is high. Preprocessing of the dataset makes sure that all the attributes (columns) are taken into account while predicting. From the above prediction and analysis, we can observe that the results obtained using Random Forest Algorithm give us an accuracy of 99% and Decision Tree Algorithm give us an accuracy of 99%. Hence this proposed method will give us an efficient method for both analysis and prediction of diabetes.

FUTURE ENHANCEMENT:

So, more early identification, detection and diagnosis is of at most importance. So, we can do prediction and analysis by using other algorithms on our dataset. The process of Feature Selection can be done more efficiently so that we can reduce the number and type of attributes. This will increase the performance of our algorithms. We can also narrow down the most important attributes or features that are useful for diabetes prediction. Healthcare professions found it hard to find healthcare data and perform analysis on them due to lack of tools, resources. But using ML, we can overcome this and can perform analysis on real-time data leading to better modelling, predictions. This enhances and improves the overall healthcare services. Now, IOT is being integrated with ML in order to make smart healthcare devices which sense if there is any change in the person's body, health data when he uses the device (Pacemaker, Stethoscope, etc.) and this will notify the person regarding this through an app. This helps in easy monitoring, advanced prediction and analysis thereby reducing errors, saving time and life of people.

REFERENCES

1. Priyanka Indoria, Yogesh Kumar Rathore (2018). A survey: Detection and Prediction of diabetes using machine learning techniques. IJERT
2. Khaleel, M.A., Pradhan, S.K., G.N Dash (2Ol3). A Survey of Data Mining Techniques on Medical Data for Finding frequent diseases. IJARCSSE.
3. K. Vembandasamy, R. Sasipriya, E. Deepa (2015). Heart Diseases Detection using Naïve Bayes Algorithm. IJISET.
4. Tawfik Saeed Zekia, Mohammad V. Malakootib, Yousef Ataeipoorc, S. Talayeh Tabibid. An Expert System for Diabetes Diagnosis. American Academic & Scholarly Research Journal Special Issue Vol. 4, No. 5, Sept 2012.
5. Vishali Bhandari and Rajeev Kumar. Comparative Analysis of Fuzzy Expert Systems for Diabetic Diagnosis. International Journal of Computer Applications (0975 – 8887) Volume 132 – No.6, December 2015.

6. Ioannis Kavakiotis, Olga Tsave, Athanasios Salifoglou, "Machine Learning and Data Mining Methods in Diabetes Research", Jan 8, 2017.

7. Eka Miranda, Edy Irwansyah, Alowisius Y. Amelga, Marco M. Maribondang, Mulyadi Salim (2016). Detection of cardiovascular Disease Risk's Level for Adults using naïve Bayes Classifier, The Korean Society of Medical informatics (KOSMI).

8. Zheng T, Xie W, Xu L, He X, Zhang Y, You M, Yang G, Chen Y (2017). A Machine Learning-Based Framework to identify Type 2 Diabetes through Electronic Health Records, International Journal of medical informatics (IJMI) Vol 9, pages 120-127.

9. Francesco Mercaldo, Vittoria Nardone, Antonella Santone (2017). Diabetes Mellitus Affected Patients Classification and Diagnosis through Machine Learning 64 Techniques, Procedia Computer Science 112 (2017) 2519-2528.

10. Rahul Joshi, Minyechil Alehegen (2017). Analysis and Prediction of Diabetes Disease using Machine Learning Algorithm: Ensemble Approach, International Research Journal of Engineering and Technology (IRJET) Volume 04 Issue 10, e-ISSN: 2395-0056.

11. Jimmy Singla, Dinesh Grover (2017). The Diagnosis of Diabetic Nephropathy using Neuro-Fuzzy expert system, Indian Journal of Science and Technology (IJST) Vol 10(28) ISSN (online) 0974-5645.

12. Mehrbakhsh Nilashi, Othman bin Ibrahim, Hossein Ahmadi, Leila Shahmoradi (2017). An Analytical method for Disease Prediction using Machine Learning Techniques, Computers and Chemical Engineering 106 (2017) 212-223.

13. Saba Bashir, Usman Qamar, Farhan Hassan Khan, Lubna Naseem (2016). HMV: A Medical Decision Support Framework using Multi-layer Classifiers for disease prediction, Journal of Computational Science 13 (2016) 10-25.

14. Mekruksavanich, S. (2016). Medical Expert System Based Ontology for Diabetes Disease Diagnosis. In Software Engineering and Service Science (ICSESS), 7th IEEE International Conference Pages 383-389, IEEE.

15. Rajeswara Rao, D., Vidyullata Pellakuri, SathishTallam, Ramya Harika, T. (2015). International Journal of Computer Science and Information Technologies (IJCSIT), Vol. 6 (2), 1103-1106