WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

An Expert System for Predicting Insulin Dosage Using Various Ml Algorithms

K. Aruna Sri[1], N. Ishwarya Jyothi[2], M. Anudeep[3], Mr. U. V. Ramesh[4] M.Tech

[1,2,3]*U. G Student, Department of CSE, Aditya Engineering College,* Surampalem, A.P., India
[4]*Assistant Professor, Department of CSE, Aditya Engineering College,* Surampalem, A.P., India

ABSTRACT

Diabetes occurs when your glucose (a type of sugar) found in the blood called as glycemia is too excess in your body. Blood glucose, which is obtained from the food you eat, serves as your body's main energy supply.

Blood glucose levels (BGLs) should appropriately balanced to allow diabetic patients to lead normal lifestyles without running the risk of long-term, serious complications.

But for a variety of factors, the majority of diabetic patients have poorly controlled blood glucose levels, which over time seriously harms the oral health, vision, hearing and mental health.

Moreover, taking the accurate amount of insulin dosage plays an important role in the treatment procedure. There are different prevention approaches like consuming strong nutriments and exercising are required to control their glycemia. In this study, we predict diabetes using the Gradient Boosting Classifier, and we predict the dosage of insulin for patients who have been identified as having diabetes using the Linear Regression Algorithm.

We are using the Pima Indians diabetes dataset and UCI insulin dosage dataset to carry out this research. With the aforementioned dataset, we are training both algorithms. Once trained, upload test dataset without class label, and Gradient Boosting will then forecast the existence of diabetes while Linear Regression forecasts the amount of insulin to administer in the event that diabetes is identified by Gradient Boosting.

1. INTRODUCTION

Diabetes mellitus, is a group of metabolic illnesses defined by excessive blood sugar levels. Diabetes is becoming further common in countries with middle- and low-income levels more rapidly.

The onset of diabetes signs can be caused by either an inadequate response to insulin or an insulin deficiency. An insulin dosage is required for a diabetic to maintain glucose control. The patient's doctor must also be aware of the necessary insulin dose based on the patient's historical records of doses as well as their present estimated records. Type 1 and type 2 diabetes are the two main subtypes of the disease. Only 5–10% of people have type 1 diabetes, whichbrought on by autoimmune cell-mediated death of the insulin-producing cells in the pancreas, which results in complete insulin insufficiency. On the other hand, 90-95% people have type 2 diabetes

The ability to anticipate glucose levels could help patients respond appropriately in critical circumstances like hypoglycemia. Therefore, a number of recent studies have looked into cutting-edge data-driven methods for creating precise predictive models of glucose metabolism. The condition in attachment to the basicinstructions that he abides by in his everyday life. Predictive representation of the glucose metabolism is the key elements of a diabetes control process.

2. LITERATURE SURVEY

Analysis of Various Data Mining Techniques to Predict the presence of Diabetes Mellitus.

AUTHORS: Devi, M. Renuka, and J. Maria Shyla

ABSTRACT: Data analysis is used to help patients' illnesses be diagnosed. Diabetes is a disease that can affect various systems in the body. The spread of diseases can be stopped and lives can be saved by early discovery. The use of various data mining techniques for early diabetes prediction is investigated in this research. The dataset used 768 cases from the Pima Indian Datafile to evaluate effectiveness in making predictions.

Predicting Diabetes Mellitus using Data Mining Techniques

AUTHORS: J. Steffi, Dr. R. Balasubramanian, Mr. K. Aravind Kumar

ABSTRACT: Diabetes is a chronic disease caused by an elevated amount of blood sugar addiction. It was explained how various automated information systems use various classifiers to predict and identify diabetes. Data analysis is used to help patients' illnesses be diagnosed. It is a condition that can impactnumerous systems in the body. The spread of diseases can be stopped and lives can be saved by early discovery. It is obvious that selecting trustworthy categories improves the system's proficiency and accuracy. Diabetes mellitus is unfairly affecting more and more households as a result of its gradually increasing prevalence. Before being diagnosed, the majority of diabetics are typically ignorant of their risk factors.

The use of data mining techniques for early diabetes prediction is investigated in this research. The dataset included 768 cases. With data from the dataset, using nine input variables and one outcome variable. The models were then evaluated.

Comparison Data Mining Techniques To Prediction Diabetes Mellitus

AUTHORS: Aswan SupriyadiSunge

ABSTRACT: Diabetes brought on by high blood sugar. Abundant self-operating algorithms are used to forecast and diagnose diabetes. One method may be used to determine the patient's illness. When predictions are present, illness can be prevented before it affects the patient, saving lives. Selecting an accurate classification increases the system's reality and accuracy as levels climb. Most diabetics aren't conscious of the dangers they were exposed to before getting sick. This method builds five predictive models with nine input factors and one output variable using the dataset's data.

3. PROBLEM STATEMENT

One of the current methods looks into advance diagnosis of diabetes applying numerous methodologies. The datafile uses 768 cases to evaluate the effectiveness in making predictions. The research shows that the Modified J48 Classifier provides the highest exactness when compared to further approaches. The dearth of and missing data in this model had an adverse effect on our precision and made it difficult to choose the feature set for the attributes.

LIMITATION OF SYSTEM

The performance of this algorithm is strongly influenced by technical issues that lead to lost or unavailable data. Additionally, it takes significantly longer to scan and organize the datasets. However, early diabetes prediction is a challenging job for medical experts due to the complicated linkage of numerous factors. Diabetes has an impact on various human systems including oral health, vision, hearing and mental health.

4. PROPOSED SYSTEM

The suggested system is modern because it uses the Gradient Boosting algorithm to predict diabetes and the Linear Regression algorithm to predict insulin dosage in patients who have been identified as having the disease. We are using the Pima Indian diabetes dataset and the UCI insulin dosage dataset to carry out this research. With the aforementioned dataset, we are training both algorithms. After training, we will upload a test dataset then Gradient Boosting will predict the presence of diabetes and Linear Regression will predict the amount of insulin to be administered in the event that diabetes is identified by Gradient Boosting. Gradient Boosting algorithm is used to forecast diabetes, and the Linear Regression algorithm is used to forecast the dosage of insulin for victims who have been identified as having diabetes. It aids individuals in taking their medicationin accurate amount.

Advantages of Proposed system:

Precise prediction is probably not possible due to the high erratic character of individual blood glucose levels and insulin dosage. It is simpler to anticipate the victim's 24-hour moderate blood glucose level and to determine whether their glucose level will be high. It makes it possible for diabetics to get the right amount of insulin. it allows diabetic patients to take care of their health.The body can function correctly and even save a person's life when insulin is given at the right time. Insulin dosages can vary significantly from person to person. A dose that one person may handle but another would deem excessive. As a result, by using this effort, we can precisely predict the insulin dosage for each person.

5. IMPLEMENTATION

5.1. Data Collection

Data collection is a procedure of collation and examining from a broad range of sources. we collect data to produce useful artificial intelligence (AI) and machine learning results, it should be collected and stored in a sense for the particular business issues.

High-performing predictive replica must be built using efficient data collection methods because predictive replicawhich they are established are good as the data. Garbage in, garbage out: the data must contain precise information that is relevant to the task at hand. For instance, rather than the number of tigers, a loan default model might benefit from rising gas costs over time.

5.2. Data Preprocessing

Data preprocessing is the method of transforming source data into something which can be utilized by machine learning algorithm. It is initial and very crucial step in the process of creating a machine learning replica. When working on a machine learning scheme, it is not always the case that we are presented with the clear and ordered data. Every time you deal with data, you must also organize and clean it up. As a result, we perform this using a data preprocessing task.

Therefore theactual data often contains errors, missing values, and may be in an unusable structure, machine learning models cannot be applied directly to it.

5.3. Training and Testing

The training dataset and the test dataset are the two core concepts in machine learning. The test dataset is helped to estimate the replica after it has created using the trained dataset. The largest (terms of size) subset of the original dataset is the trained dataset, which is used to fit the machine learning replica. Training data is first fed into the ML algorithms so they can learn how to make forecasts for the task at hand. The training material is impacted.

Unsupervised Learning Algorithms.

In unsupervised learning, the training set contains unlabeled data points because the inputs are not tagged with the associated outputs. Models must draw patterns from the given training datasets in order to make predictions.

The algorithm's accuracy and proclivity for prediction are significantly influenced by the type of training data we provide. It suggests that the model will perform better the greater the training set of data is. An ML project's complete data set, or at least 60% of it, is used for training.

It's time to evaluate the model using the test dataset after we've trained the dataset. This dataset evaluates the model's effectiveness and provides assurance that it will translate well to novel or untested datasets. A different subset of the original data from the training dataset makes up the test dataset.Information is provided for each sort of scenario the model might encounter in the real world in a well-organized dataset referred to as test data. Usually, 20–25% of the total original dataset is made up of the test dataset.

At this stage, we can also compare and contrast the accuracy of our model .If the model's accuracy on training data is greater than its accuracy on testing data, it is said to have overfitted.

The testing data should:

or some of the introductory data. It should be substantial enough to allow for precise forecasting.

If we train our model with a training set and then evaluate it with a completely different test dataset, our model won't be able to understand the correlations between the features.

.

6. MODULES/GLOSSARY

In this project, we have used many modules and everymodule has its own functionality.

6.1. Upload Diabetes Insulin Disease:

In this, we upload the user dataset which was Pima Indian diabetes dataset and UCI Insulin dosagedataset.

6.2. Execute Gradient Boosting Algorithm:

We train the dataset which we uploaded by using Gradient Boosting Algorithm.

6.3. Execute Linear Regression Algorithm:

After the dataset trained by the Gradient Boosting Algorithm then we train the dataset withLinear Regression Algorithm to predict insulin dosage amount.

6.4 Predict Diabetes & Insulin Dosage:

After the datasets are trained then it predicts the diabetes by the Gradient Boosting Algorithmand the amount of insulin dosage by the Linear Regression Algorithm.

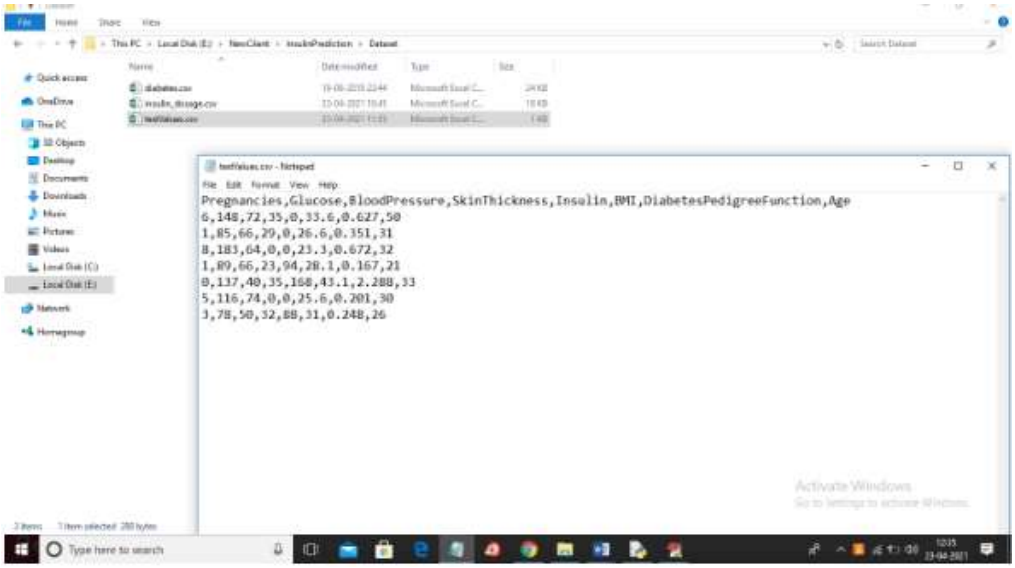
6.5. Performance Graph:

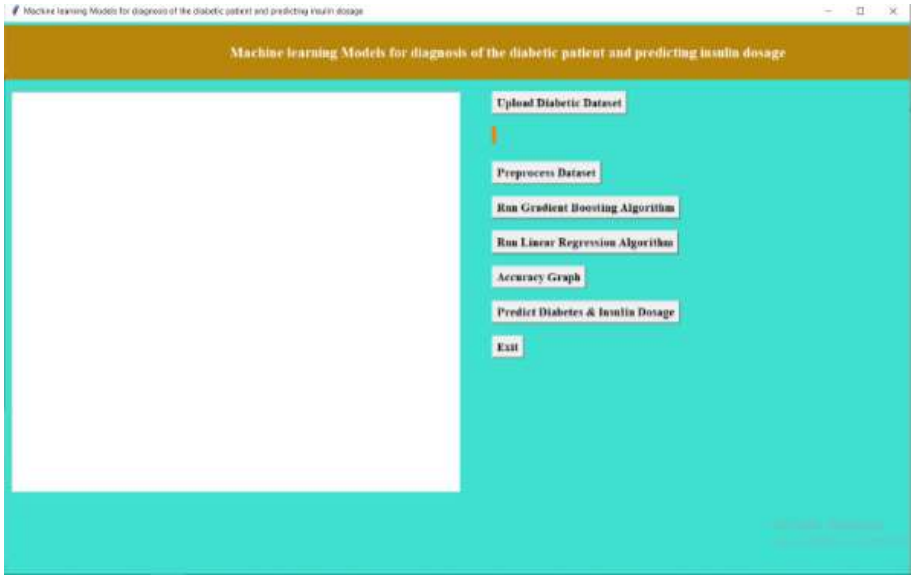
Now in this module it shows the performance graph i.e how much accuracy both the algorithmsare giving it shows in the form of graph.

7. SCREENS AND REPORTS

Models for diabetes diagnosis and dosage determination using machine learning in this research Gradient Boosting will then predict the presence of diabetes after being trained, while Linear Regression will predict the amount of insulin to administer in the case that diabetes is detected by Gradient Boosting.

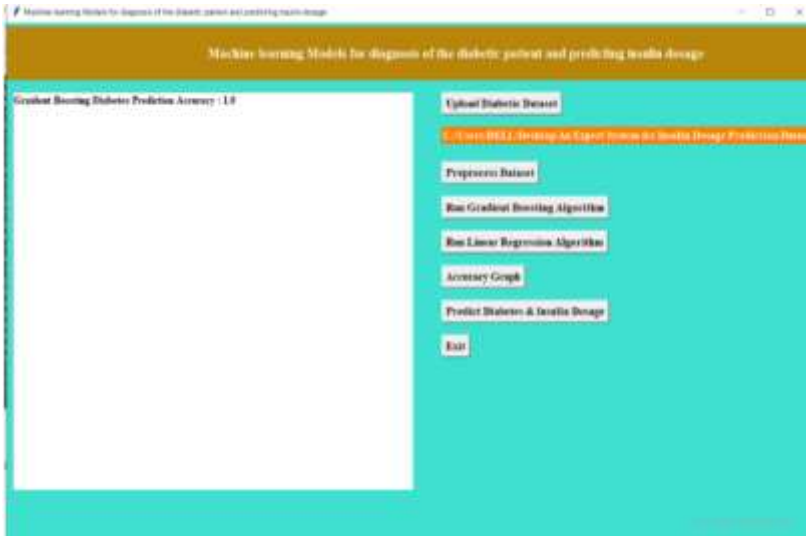
.





Dataset Loaded

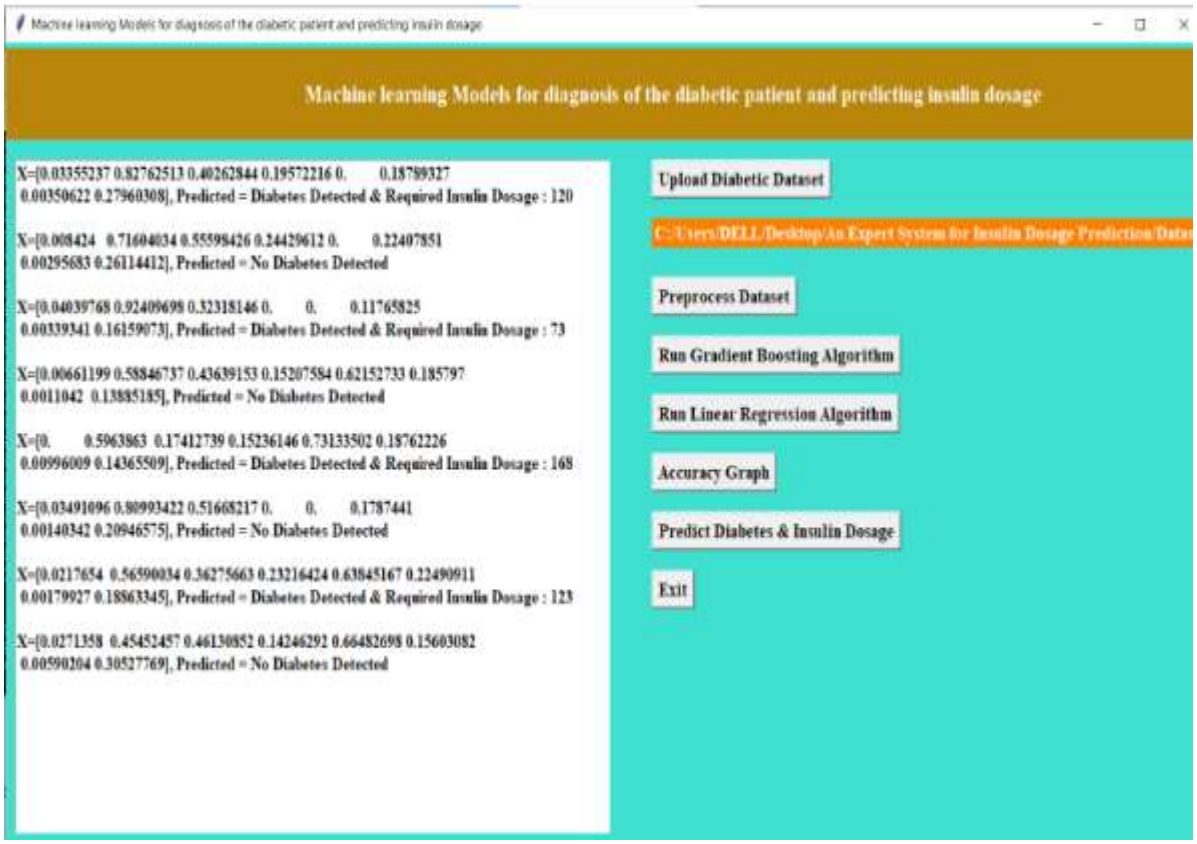
Gradient Boosting Algorithm



Linear Regression Algorithm

Accuracy Graph

Result



8. CONCLUSION

In this research, neural networks were modelled with the intention of predicting the right amount of insulin to administer to diabetic patients. The model used was a gradient boosting model trained with BP. Four patient-specific bits of information are needed by the algorithm: length, weight, blood sugar, and gender. A number of investigations used the data from 180 patients. Diabetes is predicted using the gradient boosting algorithm, and insulin dose is predicted using linear regression if diabetes is confirmed by the algorithm. The Gradient Boosting Algorithm meets results with excellent performance when compared to the Linear Regression Algorithm for the larger complex datasets.

9. FUTURE SCOPE

The research results discussed in this thesis have a lot of potential for use in a range of T1D therapy uses. The evaluation of the predictions can be moreenhanced by developing a lay of switchable replicas useful to predict Blood Glucose values This endeavor will continue with the testing of unique Blood Glucose prediction models on real subjects in a more challenging environment.

10. Bibliography

[1] American Diabetes Association. Diagnosis and Classification of Diabetes Mellitus.Diabetes Care, Vol. 31, No. 1, 2008, 55-60, 1935- 5548.

[2] I. Eleni Georga, C. Vasilios Protopappas and I. DimitriosFotiadi,."Glucose Prediction inType 1 and Type 2 Diabetic Patients Using Data Driven Techniques," Knowledge-OrientedApplications in Data Mining, InTech pp 277-296, 2011.

[3] V. Tresp,, T. Briegel, and J. Moody, "Neural-network models for the blood glucosemetabolism of a diabetic," IEEE Transactions on Neural Networks, Vol. 10, No. 5, 1999, 1204-1213, 1045-9227.

[4] C. M. Bishop, . Pattern Recognition and Machine Learning, Springer, 2006, New York.

[5] S. Haykin, Neural networks and learning machines. Pearson 2008.

[6] W.D. Patterson, Artificial Neural Networks- Theory and Applications, Prentice Hall ,Singapore. 1996.

[7] M. Pradhan and R. Sahu, "Predict the onset of diabetes disease using Artificial NeuralNetwork (ANN)," International Journal of Computer Science & Emerging Technologies, 303Volume 2, Issue 2, April 2011.

[8] W, Sandham, D,Nikoletou, D.Hamilton, K, Paterson, A. Japp and C. MacGregor, "BLOOD

Glucose Prediction for Diabetes THERAPY USING A RECURRENT Artificial NeuralNetworks ", EUSIPCO, Rhodes, 1998, PP. 673-676

[9] M, Divya, R. Chhabra, S. Kaur and S. ,"Diabetes Detection Using Artificial NeuralNetworks & Back-Propagation Algorithm", International Journal of Scientific & TechnologyResearch". V 2, ISSUE 1, JANUARY 2013.

[10] G, Robertson, E. Lehmann, W. Sandham, and D. Hamilton Blood, "Glucose PredictionUsing Artificial Neural Networks Trained with the AIDA Diabetes Simulator: A Proof-of-Concept Pilot Study". Journal of Electrical and Computer Engineering , V 2011 (2011), ArticleID 681786, 11 pages.