International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

A EXPERT SYSTEM FOR PREDICTING INSULIN DOSAGE USING VARIOUS ML ALGORITHMS

Ms.M.ANITHA[1], Mr. CH. SATYANARAYANA REDDY [2], Ms. K. YAMINI[3]

#1 Assistant professor in the Master of Computer Applications in the SRK Institute of Technology, Enikepadu, Vijayawada, NTR District

#2 Assistant professor in the Master of Computer Applications SRK Institute of Technology, Enikepadu, Vijayawada, NTR District

#3 MCA student in the Master of Computer Applications at SRK Institute of Technology, Enikepadu, Vijayawada, NTR District

ABSTRACT: You can acquire diabetes if your blood glucose, also referred to as blood sugar, is too high. Blood glucose, which is obtained from the food you eat, serves as your body's main energy supply. Long-term metabolic disease is diabetes mellitus. Blood glucose levels (BGLs) should be appropriately adjusted 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 heart, blood vessels, eyes, kidneys, and nerves.

However, taking the correct quantity of insulin dosage has the crucial role in the treatment process. Traditional prevention techniques like eating healthy food and exercising are necessary for diabetic patients to control their BGLs. 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 diabetes dataset and the UCI insulin dosage dataset to carry out this research. With the aforementioned dataset, we are training both algorithms. Once trained, we will upload a test dataset without a class label, and Gradient Boosting will then predict the presence of diabetes while Linear Regression will predict the amount of insulin to administer in the event that diabetes is identified by Gradient Boosting..

1. INTRODUCTION

Diabetes, also referred to as diabetes mellitus, is a group of metabolic illnesses defined by high blood sugar levels brought on by deficiencies in insulin secretion, action, or both. Diabetes is becoming more common in countries with middle- and low-income levels more rapidly. People with diabetes increased from 108 million in 1980 to 422 million in 2014. Globally, the prevalence of diabetes among adults over 18 rose from 4.7% in 1980 to 8.5% in 2014.

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 blood sugar level. Type 1 and type 2 diabetes are the two main subtypes of the disease. Only 5–10% of people with diabetes have type 1 diabetes, which is brought on by autoimmune cell-mediated death of the insulin-producing cells in the pancreas, which results in complete insulin deficiency.. On the other hand, type 2 diabetes is a more prevalent category (i.e., accounts for ~90-95% of those with diabetes) and is a combination of resistance to insulin action and an inadequate compensatory insulin secretion.

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. Numerous diabetes management systems have been proposed to aid the patient in the self-management of the condition in addition to the general guidelines that he abides by in his everyday life. Predictive modeling of the glucose metabolism is one of the key elements of a diabetes control system..

2. LITERATURE SURVEY

Water demand forecasting using extreme learning machines

AUTHORS: Tiwari, Mukesh, Jan Adamowski, and Kazimierz Adamowski

ABSTRACT: Extreme learning machine (ELM) modeling approaches that were recently created were assessed and compared to comparable conventional artificial neural network-based models for daily urban water demand. (i.e., ELM, ELMW, and ELMB). (i.e., ANN, ANNW, ANNB). The urban water demand forecasting models were developed using the three-year water demand and climate records for Calgary, Alberta, Canada. The hybrid ELMB and ANNB models generated acceptable 1-day lead-time forecasts with comparable accuracy, whereas the ANNW and ELMW models offered superior accuracy, with the ELMW model outperforming the ANNW model. A significant improvement in peak urban water demand prediction was only made feasible with the ELMW model. The success of the ELMW model over the ANNW and ANNB models demonstrated how crucial wavelet transformation is to improving the overall efficacy of the urban water demand model.

Analysis of Various Data Mining Techniques to Predict Diabetes Mellitus.

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

ABSTRACT: Data analysis is used to help patients' illnesses be diagnosed. Diabetes mellitus is a chronic condition 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 Dataset to evaluate the effectiveness of data mining techniques in making predictions. Analysis shows that Modified J48 Classifier provides more accuracy than alternative techniques..

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. Diabetes mellitus is a chronic condition that can affect various 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 from the PIMA Indian Diabetes Dataset in order to evaluate the effectiveness of data mining techniques in prediction. After that, we built five predictive models with data from the dataset, using nine input variables and one outcome variable. The accuracy, precision, sensitivity, specificity, and F1 Score metrics of the five models were then evaluated. This study compares the performance analysis of Nave Bayes, Logistic Regression, Artificial Neural Networks (ANNs), C5.0 Decision Tree, and Support Vector Machine (SVM) models to forecast diabetes using common risk factors. The decision tree model (C5.0) had given the best classification accuracy, followed by the logistic regression model, Naïve Bayes, ANN and the SVM gave the lowest accuracy IndexTerms—Data mining, Prediction, Naïve Bayes, Logistic Regression, C5.0 Decision Tree, Artificial Neural Networks (ANN) and Support Vector Machine (SVM).

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

Comparison Data Mining Techniques To Prediction Diabetes Mellitus

AUTHORS: Aswan Supriyadi Sunge

ABSTRACT: Diabetes is one of the chronic diseases brought on by high blood sugar. Various automated algorithms are used to forecast and diagnose diabetes. One data mining 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. In this study, the efficacy of Naive Bayes, Decision Trees, SVMs, K-NNs, and ANNs were examined to predict diabetes mellitus.

3. PROBLEM STATEMENT

One of the current methods looks into the early diagnosis of diabetes using various data mining techniques. The dataset used 768 cases from the PIMA Indian Dataset to evaluate the effectiveness of data mining techniques in making predictions. The research shows that the Modified J48 Classifier provides the highest accuracy when compared to other 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 task for medical experts due to the complex interdependence of numerous factors. Diabetes has an impact on various human systems, including the kidney, eye, heart, nerves, foot, and others.

.

4. PROPOSED SYSTEM

The suggested system is modern because it uses the Gradient Boosting Classifier to predict diabetes and the Logistic Regression algorithm to predict insulin dosage in patients who have been identified as having the disease. We are using the PIMA 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 without a class label, at which point 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 Classifier is used to forecast diabetes, and the Linear Regression algorithm is used to predict the dosage of insulin for patients who have been identified as having diabetes. It aids individuals in taking their medication. in 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 patient's 24-hour average 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. By using the Gradient Boosting Classifier to predict diabetes and the Linear Regression algorithm to predict insulin dosage in diabetic detected patients, 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 the procedure of gathering and examining information from a broad range of sources. To use the data we collect to produce useful artificial intelligence (AI) and machine learning solutions, it must be collected and stored in a way that makes sense for the particular business issue at hand.

The collection of data allows you to maintain a record of earlier occurrences so that you can use data analysis to find recurring patterns. From these patterns, you develop predictive models using machine learning algorithms to look for trends and anticipate upcoming changes.

High-performing predictive models must be built using efficient data collection methods because predictive models are only as good as the data on which they are founded. 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 process of transforming raw data into something that can be used by a machine learning algorithm. It is the first and most crucial step in the process of creating a machine learning model. When working on a machine learning project, 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.

.Since real-world data frequently contains noise, missing values, and may be in an unusable structure, machine learning models cannot be applied directly to it. The accuracy and efficiency of a machine learning model are increased through data preprocessing, which is required to clear the data and prepare it for the model.

5.3.Training and Testing

The training dataset and the test dataset are the two core concepts in machine learning. The test dataset is used to evaluate the model after it has been created using the training dataset. The largest (in terms of size) subset of the original dataset is the training dataset, which is used to train or fit the machine learning model. 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 by the use of either supervised learning or unsupervised learning.

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. For supervised learning, on the other hand, names are included in the training data to aid in the model's training and forecasting.

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 it using the training 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. It serves as a benchmark once model training is complete because it has some similar features and a similar class chance distribution. 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 when applied to the test dataset with the accuracy of our model when applied to the training dataset. 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

International Journal For Advanced Research
In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

understand the correlations between the features.

.

6. MODULES/GLOSSARY

A Module is a collection of source file and build settings that allow you to divide your project into discrete units of functionality. In our project, we have used many modules and every module has its own functionality.

6.1. Upload Diabetes Insulin Disease:

In this, we upload the user dataset which was PIMA diabetes dataset and UCI Insulin dosage dataset.

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 with Linear 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 Algorithm and the amount of insulin dosage by the Linear Regression Algorithm.

6.5. Performance Graph:

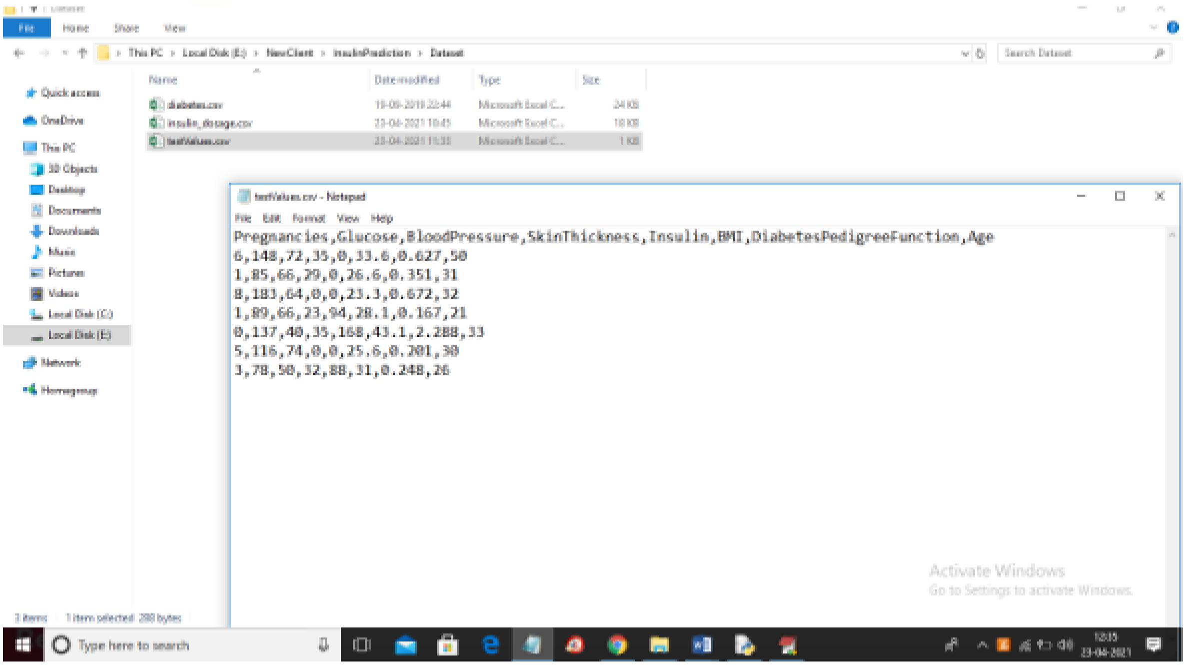
Now in this module it shows the performance graph i.e how much accuracy both the algorithms are giving it shows in the form of graph.

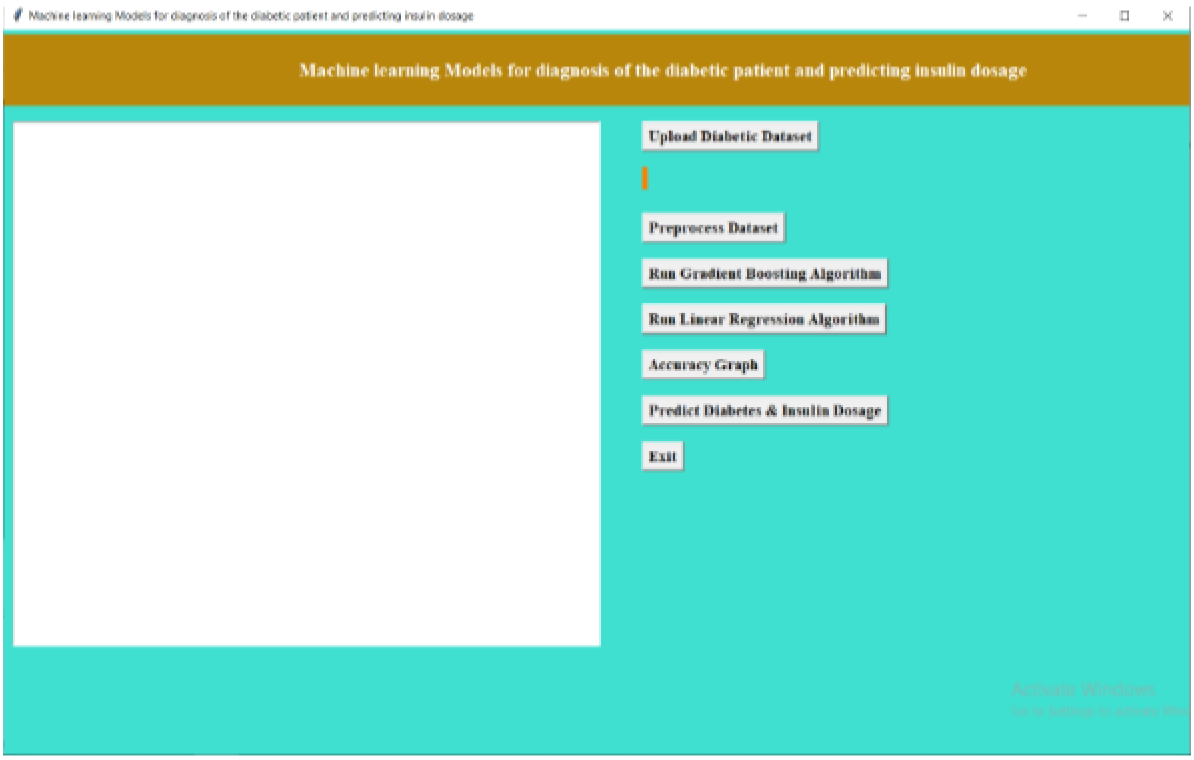
7.SCREENS AND REPORTS

Models for diabetes diagnosis and dosage determination using machine learning In this research, we use the Gradient Boosting Classifier to predict diabetes and the Linear Regression algorithm to predict the insulin dosage for patients who have been diagnosed with diabetes. To conduct this study, we are utilizing the PIMA diabetes dataset and the UCI insulin dose dataset. We are training both algorithms on the aforementioned dataset. 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.

.

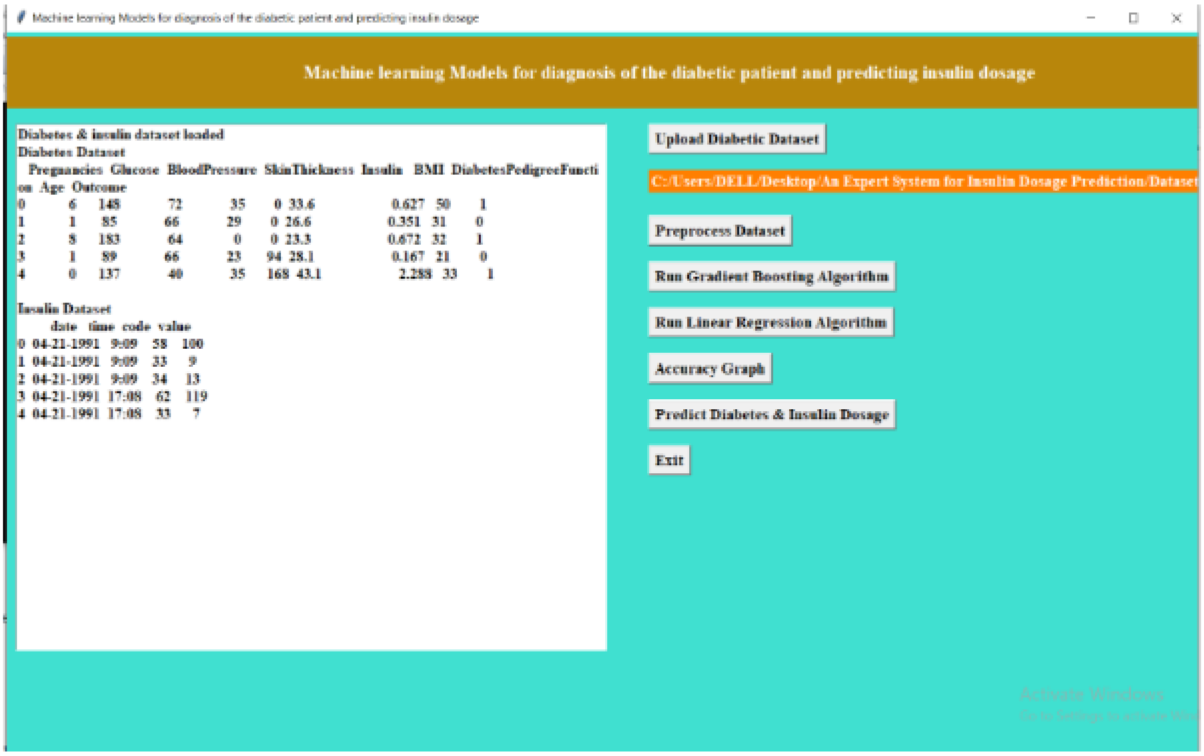
International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362
IJARST



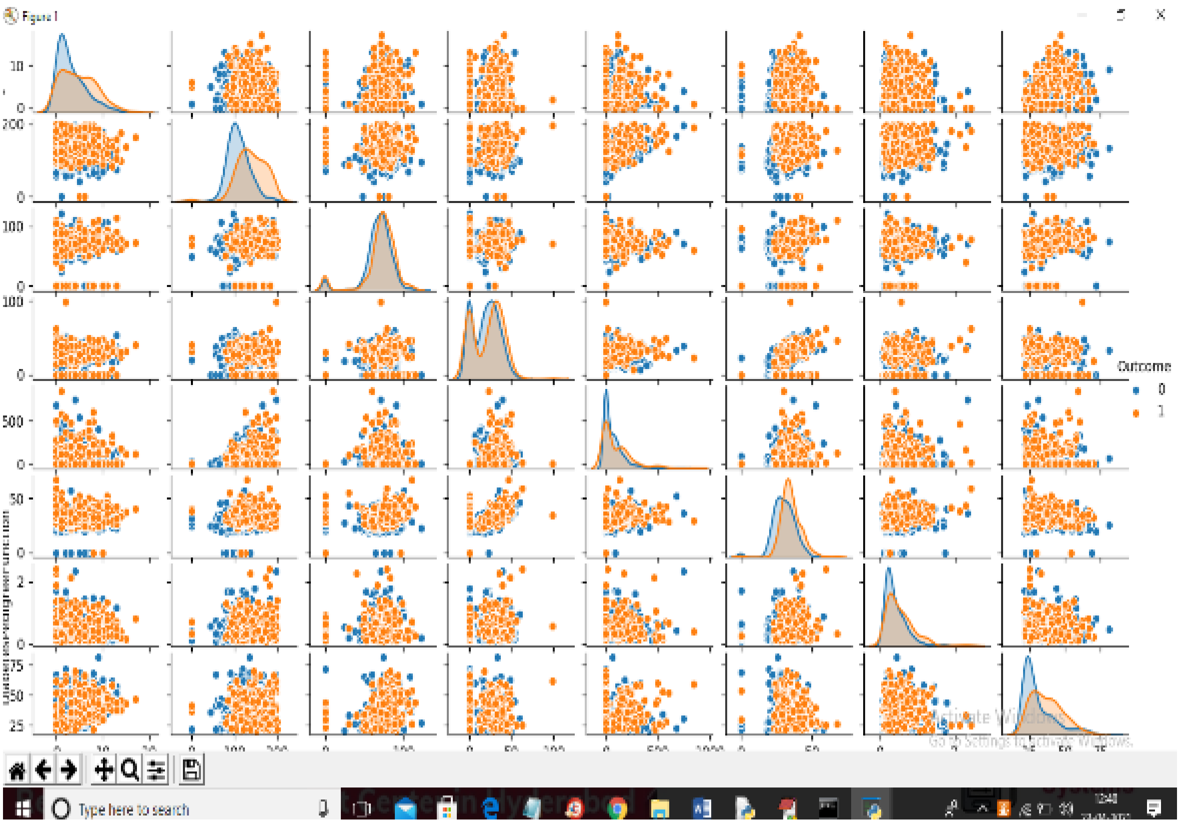


Dataset Loaded

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362



In above screen we can see both datasets loaded and we can see some records from each dataset and will get below graph also

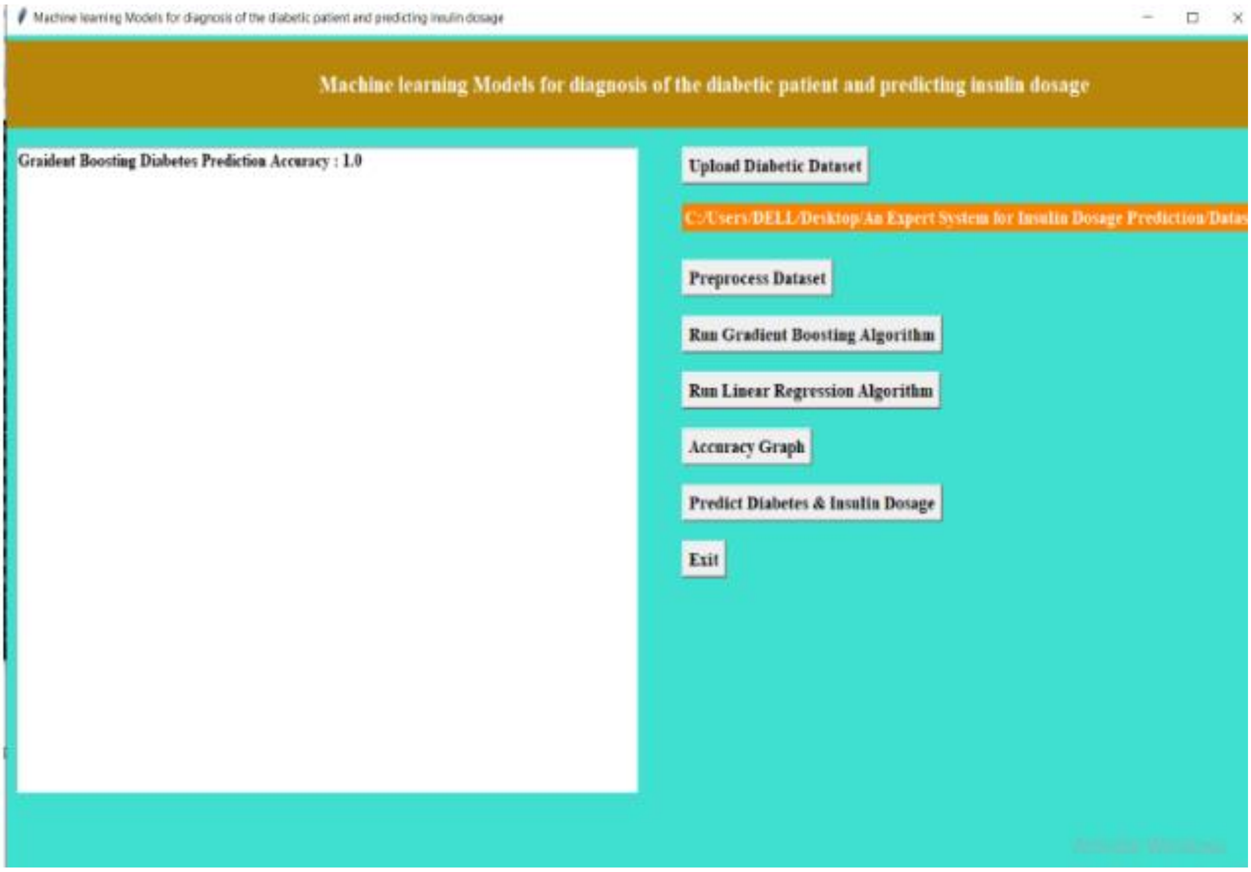


In above graph we can see diabetes for each column where red colour dots indicate presence of diabetes and blue represents no diabetes detected.

We are plotting graph for each column value to show with which value diabetes is present and with which value diabetes is not present for example in above graph in first column we are plotting graph for ‘number of Pregnancies’ with ‘presence or no presence of diabetes’ and now close above graph and then click on ‘Preprocess Dataset’ button to remove missing values and to split dataset into train and test.

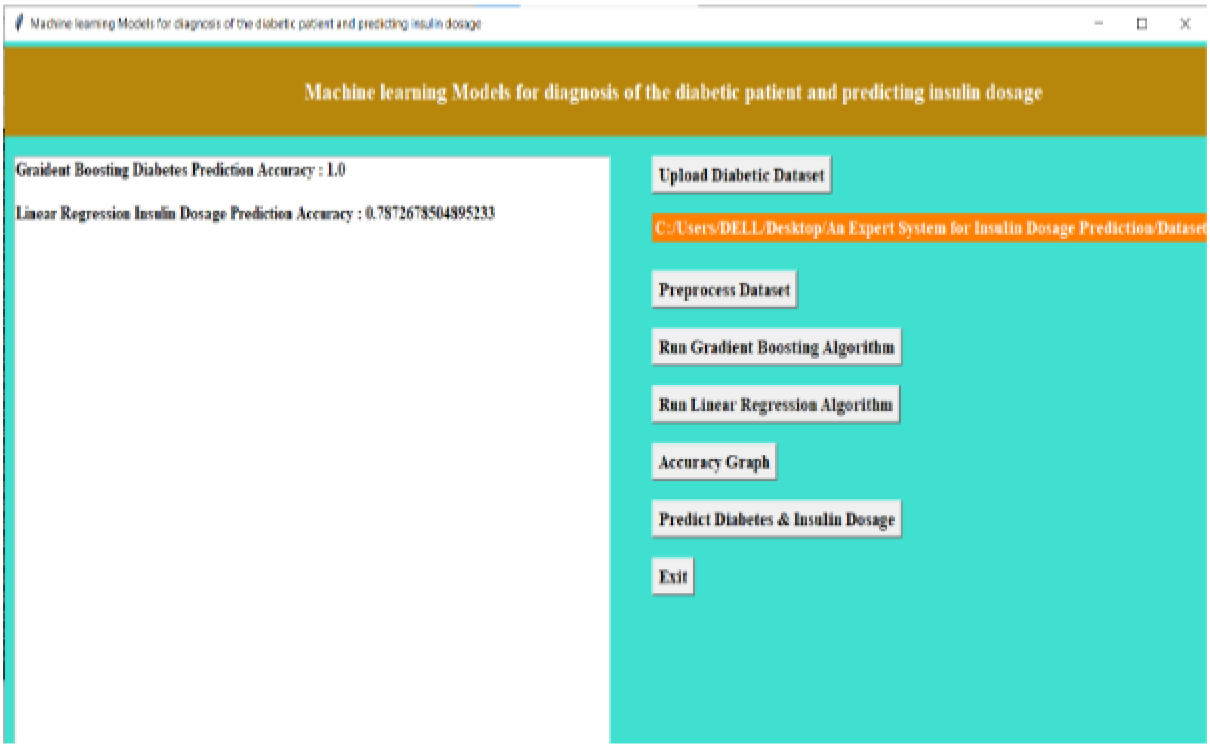
Gradient Boosting Algorithm

Now click on ‘Run Gradient Boosting Algorithm’ button to train gradient boosting with above dataset to predict the diabetes in the patients.



Linear Regression Algorithm

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362



Accuracy Graph

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
IJARST
ISSN: 2457-0362

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 converged quickly and generated results with excellent performance when compared to the Linear Regression Algorithm.

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 performance and safety of the predictions can be further improved by developing a set of interchangeable models that predict useful Blood Glucose values for control and therapy purposes based on the identification of individual specific dynamics, lifestyle, and other factors. This endeavor will continue with the testing of unique Blood Glucose prediction models on real subjects in a more challenging environment.

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
IJARST
ISSN: 2457-0362

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. Dimitrios Fotiadi,. "Glucose Prediction in Type 1 and Type 2 Diabetic Patients Using Data Driven Techniques," Knowledge-Oriented Applications in Data Mining, InTech pp 277-296, 2011.

[3] V. Tresp,, T. Briegel, and J. Moody, "Neural-network models for the blood glucose metabolism 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 Neural Network (ANN)," International Journal of Computer Science & Emerging Technologies, 303 Volume 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 Neural Networks ", EUSIPCO, Rhodes, 1998, PP. 673-676

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

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

AUTHOR PROFILES

Ms. M. ANITHA completed her Master of Computer Applications and Masters of Technology. Currently working as an Assistant professor in the Department of Master of Computer Applications in the SRK Institute of Technology, Enikepadu, Vijayawada, NTR District. Her area of interest

includes Machine Learning with Python and DBMS.

Mr. Ch. SATYANARAYANA REDDY Completed his Bachelor of Computer Applications at Acharya Nagarjuna University. He completed his Master of Computer Applications at Acharya Nagarjuna University. Currently working as an Assistant professor in the Department of Computer Applications SRK Institute of Technology, Enikepadu, Vijayawada, NTR(DT). His areas of interest include Networks, Machine Learning & Artificial Intelligence.

Ms. K. YAMINI is an MCA student in the Department of Computer Applications at SRK Institute of Technology, Enikepadu, Vijayawada, NTR District. She has a Completed Degree in B.Sc.(computers) from Sri Durga Malleswara Siddhartha Mahila Kalasala at MG road. Her areas of interest are DBMS and Machine Learning with Python.