PERFORMANCE ANALYSIS OF DIABETES PREDICTION USING MACHINE LEARNING MODELS

T.Vivek , G.Yadhu Vamsi, P.Susanna, **A.Apoorva, M.Lahari, Mrs.R.Umamaheswari (HOD of Dept)**, Department Of Electronics And Computer Engineering, Vignan's Institute of Information Technology, Visakhapatnam-India

Abstract

The development of diabetes is a global health crisis that can affect the lifes of millions . It is a disease that effect when the level of glucose of blood increases. It can cause various symptoms such as increased hunger and thirst, kidney failure, and blindness. In the past, patients were advised to consult a doctor if they had diabetes. By the help of ML, we were able to develop a resilient system for treating this condition. ML is a scientific discipline that involves learning from experience. This paper aims to design a system that can predict the diabetic cause level, of a person with high accuracy. The system is built on three main algorithms: the random forest, the logistic regression, and the decision tree classifier.

Keywords—: Diabetes, ML,RF, LR, DT classifier, Python Programming

I. INTRODUCTION

Diabetes is a long-continuing complaint in which the position of blood glucose is above normal.The food consumed is converted into glucose or sugar for our bodies to use for energy.The pancreas present in the gut releases a hormone is called insulin to help glucose enter into cells of the body either doesn't induce enough insulin or can't use its own insulin as well as it should.this results in erecting up of sugar situations in the blood.The high position of blood glucose affects multiple organs of mortal body.The major signs and symptoms of diabetes are extreme thrist,increased urination,unexplained hunger,blurred vision,frequent infections.indeed though there's no cure for diabetes,it must be treated and controlled in the early stage.Diabetes is farther divided in to two types type-1 is IDDM(insulin dependent diabetes mellitus) and type-2 is NIDDM(non insulin dependent diabetes mellitus). The type-1 diabetes issues due to the pancreas failure to supply hypoglycemic agent.This type one is substantially plant in childs below twenty times old. People faces throughout their life because of the type-1 diabetic and stop insulin vaccinations.The type-2 diabetes starts with hypoglycemic agent resistance.The sickness develops because of absence of hypoglycemic agent that also erected.The people affected by type-2 diabetes will be high by 2025. The rotundity is one of the main reasons for type two diabetes.Early vaccination is the key in diabetes because early vaccination and proper treatment can help health complications.the styles have been proposed and published for diabetes vaccination using machine learning.In the exploration paper,diabetes is prognosticated using colourful significant attributes and the relationship between them is also characterized algorithms employed are arbitrary number,logistic regression and decision tree.The Main study of bracket systems give high delicious with high running time and low perfection indeed with enormous dataset.currently,numerous effective analysis ways are there at effective.data analysis change discovery in ultra modern hospitals.

II. RELATED WORKS

In Literature,Priyanka sonar et.al[1] have Jayamalini have considered that diabetes is the disease which is occurred due to production of insufficient amount of insulin in the blood.various machine learning algorithms were utilized for producing accurate results.presently,multitudinous effective analysis ways are there for effective cost.these data analysis changes discovery in modern hospitals.

MD.Faisal Faruque et.al[2]made the performance comparison of different machine learning ways and estimated the vacination results grounded on the applicable threat factors of diabetes mellitus. SVM, NB, KNN, DT and employed. MD.Kamrul Hasan et.al[3] proposed a frame for the vatication of diabetes using data standardization, point selection,k-foldcross-validation and colorful machine learning classifiers. Taiyu zhu et.al[4] have proposed a new deep RL algorithm that optimises rudimentary insulin and glucagon delivery using generalised DQN training. Mohammed Azeem Sarwar et.al[5] in their study used various datamining tools and techniques to convert raw data into meaningful information by computers.Supervised and Unsupervised and Reinforcement Machine learning algorithms are implemented .Soumen Moulik et.al[6] analysed colorful supervised machine literacy ways like KNN, NB, DT Classifiers, RF and SVM. Among them KNN redounded in loftiest delicacy.Nikos Fazakis et.al[7] examined several algorithms similar as Naïve Bayes, LR, ANN, SVM, RF that integrates the long term T2DM Threat Vacination Model.Juncheng Ma et.al[8]Studied six machine literacy algorithms to prognosticate the diabetes in early stage where three algorithms arbitrary timber, boosting, neural networks are having high delicacy.Koouthar Driss et.al[9] had analysed K nearest neighbour to work on two criterias F1 score, ROC and better accuracy is achieved through KNN algorithm.T.Jayalakshmi et.al[10] Studied the impact of Preprocessing , missing values ,applied propagation algorithm and missing data analysis where the accuracy was tremendously proved by the using the combination of missing values and Preprocessing.Taiyu Zhu et.al[11] used the deep learning algorithms to predict the diabetes. This resulted in superior experimental performance than other approaches.A.Mary Posonia et.al[12] applied Decision tree j48 classifier to predict diabetes and it produced 91.2% accurate results by using decision trees.G.A.Pethunachaiyar et.al[13] examined diabetes vaccination using kernel support vector machines used other kernel functions like LKF, PKF, RKF. SVM with LKF gives 100 results.Tannu Chauhan et.al[14] used supervised and unsupervised machine learning for diabetes. Decision tree classifier have the potential to detect the diabetes in early stage and combination of PCA and K-Means produced good results.Lejla Alic et.al[15] has analysed Linear svm exercising the MATLAB'S SVMT rain function that showed big values of glucose observed at 2h mark during the OGTT that explosively indicates the implicit threat of future.

Many Researchers works on so many machine learning Algorithms but no one taken combination of RF,DT,LR to predict accuracy for diabetes prediction

In this paper we are calculating accuracy for the 3 algorithms and gives accurate diabetes prediction.for that we are taking datset and then python programming and get the results.

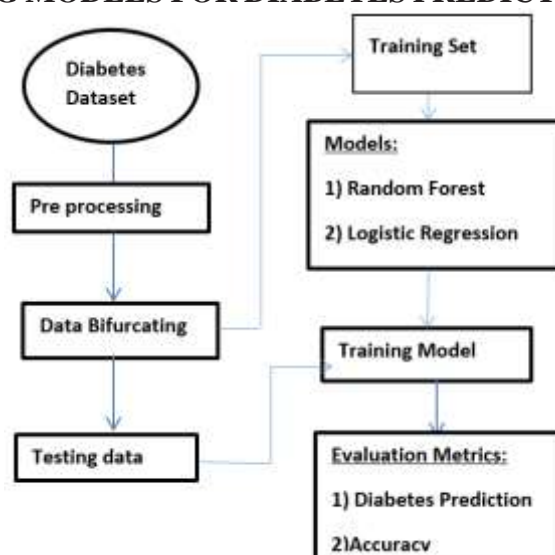## II. MACHINE LEARNING MODELS FOR DIABETES PREDICTION

Fig 2 Block Diagram For Diabetes Prediction using ML Models

Accuracy

Accuracy is a statistic that sums up how well a model performs across all classes. It's helpful when all of the classes are equally important. The ratio between the number of right guesses and the total number of forecasts is used to compute it.

Accuracy = (True Positive + True Negative)/(True Positive + False Positive +True Negative + False Negative)

III. ANALYSIS USING LOGISTIC REGRESSION :

Logistic regression generally the probability which lies between 0 or 1.the probability to classifies the data if it's lesser than 0.5 it belongs one order lower than 0.5 belongs to another order.we pass the weighted sum of inputs through an activation function that can collide values in between 0 or 1.similar AF is the sigmoid function is SF wind attained is S-Wind sigmoid function.The independent variables should be independent of each other model should have little or no collinearity.
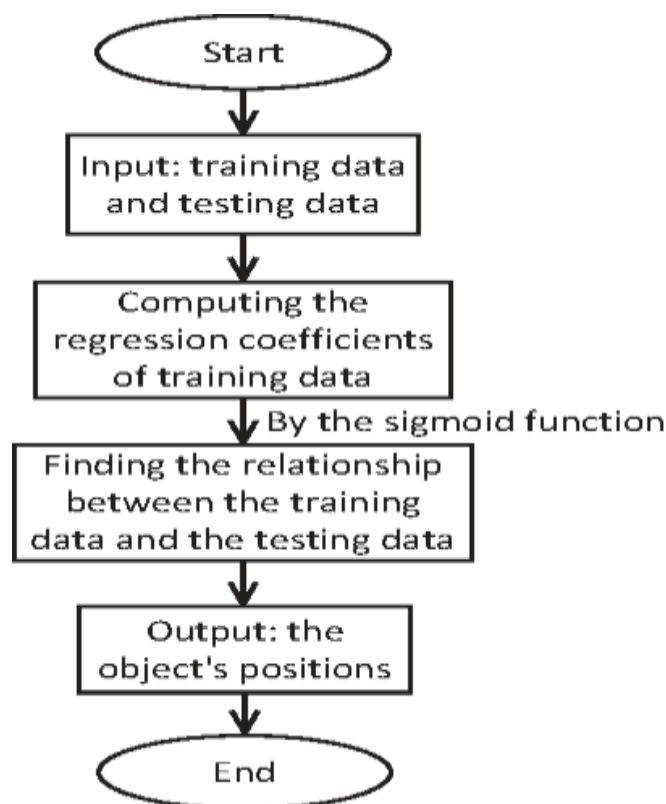

Fig. 3. Flowchart of Logistic Regression

By using this commands in the python program:

```
lr_model = LogisticRegression(random_state=1)
lr_model.fit(x_train, y_train)
lr_predict = lr_model.predict(x_test)
```

For accuracy this are the commands

```
lr_acc = accuracy_score(y_test,lr_predict)
acc_data.append(lr_acc)
print("\n* Logistic Regression Report")
print(" - accuracy score: ",round(lr_acc,4))
```

IV. ANALYSIS USING DECISION TREE CLASSIFIER:

Decision tree is to supply model that predicts the worth of a target variable,for which the choice tree uses the tree representation to interrupt problem.it typically suffer from the matter of overfitting if it's allowed to grow with none control.when a knowledge set with features is taken as input by a choice tree it'll devise some set of rules to de vaccination.Mathematics behind decision tree algorithm is first we've find entropy is nothing we've standard formula.point values are preferred to be categorical,the values continue also they're converted as separate before erecting the model.
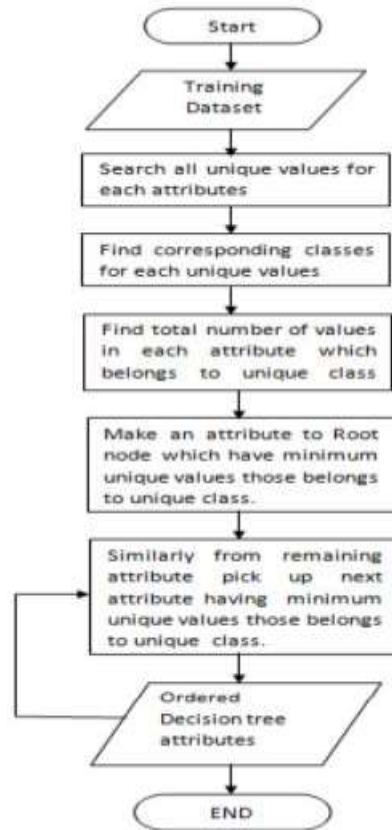

Fig. 4. Flowchart of Decision tree algorithm

By using this commands for decision tree algorithm
Dt_model = tree.DecisionTreeClassifier()
Dt_model.fit(x_train, y_train)
Dt_predict = Dt_model.predict(x_test)

The accuracy commands are

Dt_acc = accuracy_score(y_test,Dt_predict)
acc_data.append(Dt_acc)
print("\n* Decision Tree Report")
print(" - accuracy score: ",round(Dt_acc,4))

V. ANALYSIS USING RANDOM FOREST ALGORITHM:

RF builds DT on different samples and their maturity v for bracket and normal just in case of regression.created from the subsets of data and the last affair is grounded on mean or maturity ranking and hence overfitting problem is taken care.occasionally arbitrary number is slower than the decision tree it can maintain the data set containing nonstop variables in case of regression and categorical variables in case of classification.it performs better for the bracket problems.
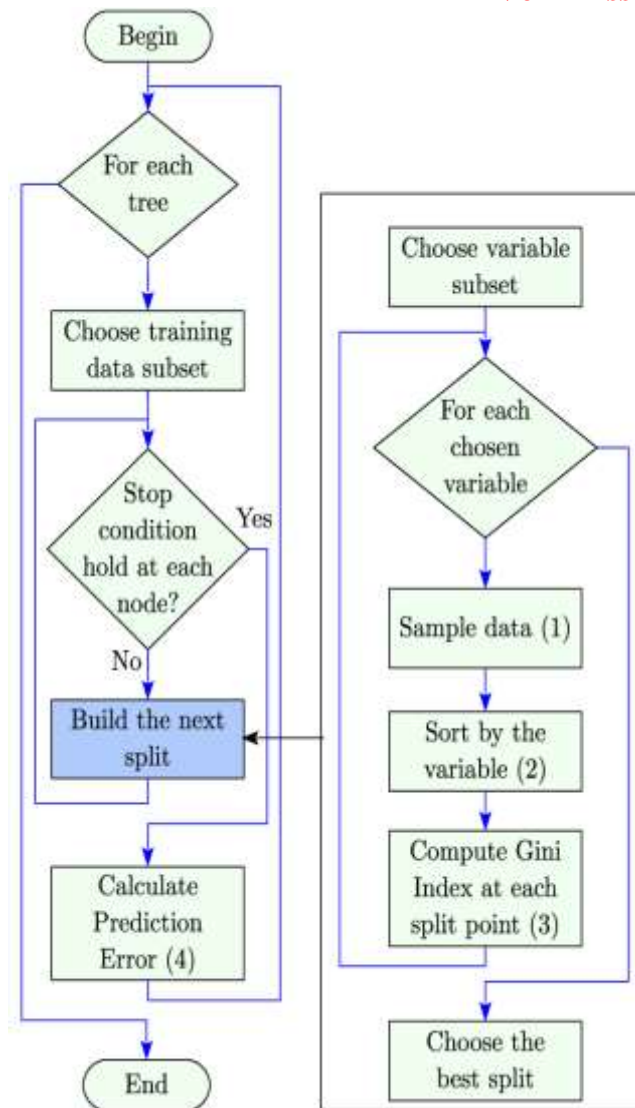
Fig. 5. Flowchart of Random Forest.

By using the commands for random forest

rfc_model = RandomForestClassifier(random_state=1)
rfc_model.fit(x_train, y_train)
rfc_predict = rfc_model.predict(x_test)

the accuracy commands are

rfc_acc = accuracy_score(y_test,rfc_predict)
acc_data.append(rfc_acc)
print("* Random Forest Classification Report")
print(" - accuracy score: ",round(rfc_acc,4))

VI. COMPARATIVE ANALYSIS :

ALGORITHM	ACCURACY (%)-1	Accuracy(%)-2
Random Forest	81.17%	96.25%

Logistic Regression	77.92%	78.75%
Decision Tree	70.13%	94.56%

Table 1. Comparing results of RF,LR & DT

Hence, the highest accuracy obtained in this analysis is 81.7% and 96.25% by random forest algorithm. It is clearly observed from the above table that the random forest algorithm has given best results in terms of accuracy.
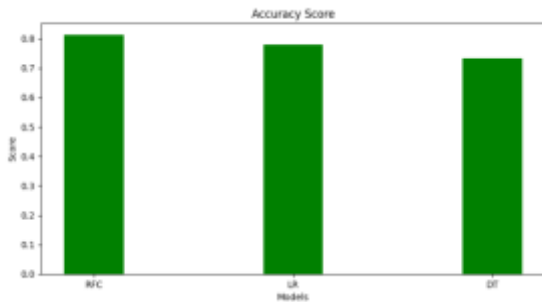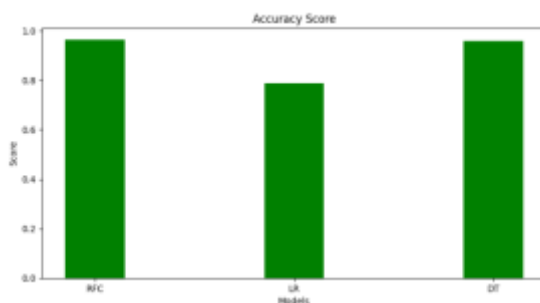

Fig 6 Bar Graph For Accuracy-1

Fig 7 Bar Graph for Accuracy-2

VII. CONCLUSION:

Accuracy of random forest is significantly higher than logistic regression but comparison higher in case of DT.This is because the RF is a collection of DT as we know the decision tree obtains its accuracy through its mathematical approach by categorising the data into groups.In random forest it is enhanced down to the undivided field of the data which increases the accuracy when compared to decision tree

REFERENCES

[1] Veena Vijayan V. and Anjali C, Prediction and diagnosis of Diabetes Mellitus, "A Machine Learning Approach",2015 IEEE recent Advance in Intelligent Computational Systems.

[2] Aakansha Rathore and Simran Chauhan " Detecting and Predicting Diabetes Using Supervised Learning". International journal of advancement Research in Computer Science, Volume:08,May-June 2017.

[3] MD.Kamrul Hassan and Eklas Hossain"Diabetes prediction Using Ensembling Of Different Machine Learning Classifiers"Published in 2107.

[4]. Priyanka Sonar and K.JayaMalini "Diabetes forecast Using ML Approaches" Proceedings of the 3rd Conference Volume:10 June-July 2019.

[5] Taiyu Zhu, Kezhi Li and Pau Herrero "Basal Glucose Control in Type-1 Diabetes Using Reinforcement Learning : An In Silico Validation"Published in 2020.

[6] Md.Faisal Faruque and Asaduzzaman "Performance Analysis Of Machine Learning Techniques To Predict Diabetes Mellitus"2019 international conference on Electrical ,Computer and communication Engineering(ECCE),7-9 February,2019.

[7] Muhammad Azeen Sarwar, Nasir Kamal and Wajeeha Hamid "Prediction Of Diabetes Using Machine Learning Algorithms in Healthcare" Published in the year 2018.

[8] Arwatki Chen Lyngdoh, Nurul Amin Choudhury and Soumen Moulik "Diabetes Prediction Using Machine Learning Algorithms" 2018 IEEE EMBS Conference on Biomedical Engineering and Sciences(IECBES).

[9] Arwatki Chen Lyngdoh, Nurul Amin Choudhury and Soumen Moulik "Diabetes Prediction Using Machine Learning Algorithms" 2018 IEEE EMBS Conference on Biomedical Engineering and Sciences(IECBES).

[10]. Nikos Fazakis and Otilia Kocsis "Machine Learning Tools for Long Term Type-2 Diabetes Risk Prediction"Published July 20,2021.

[11] Nikos Fazakis and Otilia Kocsis "Machine Learning Tools for Long Term Type-2 Diabetes Risk Prediction"Published July 20,2021.

[12] Juncheng Ma "Machine Learning in Predicting Diabetes in the Early Stage" Published in the year 2020.

[13] A.Mary Posonia and S.Vigneshwari "Machine Learning Based Diabetes Prediction using Decision Tree J48" Published in the year 2020.

[14] Tannu Chauhan and Samrath Malik "Supervised and Unsupervised Machine Learning based Review on Diabetes Care"2021 7th International Conference on advance computing & communication systems(ICACCS)

[15] Lejla Alic and Hasan T Abbas "Predicting Diabetes in Healthy Population through Machine Learning" 2019 IEEE 32nd International Symposium on Computer Based Medical system(CBMS)