IJCSE International Journal of Computer Sciences and Engineering Open Access

Research Paper Vol. 9, Issue.9, September 2021 E-ISSN: 2347-2693

Diabetic Disease Prediction System using Supervised Machine Learning Approaches

Ommi Ramu[1*], Brahmaji Godi[2], Om Prakash Samantray[3]

[1,2,3]Department of Computer Science and Engineering, Raghu Institute of Technology (Autonomous), Affiliated to JNTU (K), Visakhapatnam, Andhra Pradesh, India

[*]Corresponding Author: brahmajigodi@gmail.com, Tel.: +91 8500925485

DOI: https://doi.org/10.26438/ijcse/v9i9.7582 | Available online at: www.ijcseonline.org

Received: 18/Sept/2021, Accepted: 20/Sept/2021, Published: 30/Sept/2021

***Abstract*—**In the present study Diabetics is one of the critical diseases which can fall at any group of age and gender. The major causes lead to diabetics is mostly inheritance, in a proper healthy lifestyle, Irregular food habits, stress, and no physical exercise. Prediction of Diabetics is a very important study since it is one of the leading causes of sudden kidney failures, heart attacks, and brain stroke etc. The diabetic patient treatment can be done through patient health history. The Doctor can find hidden information about the patient through healthcare applications and it will be used for effective decision-making for the patient’s health condition. The healthcare industry is also collecting a large amounts of patient health information from different data warehouses. Using these healthcare databases researchers used to extract information for predicting the diabetics of the patient. Researchers are focused on developing software with the help of machine learning methods that can help clinicians to make better decisions about a patient's health based on their prediction and diagnosis. The main purpose of this program is to diagnose a patient's diabetes using machine learning methods. A relative study of the various competences of machine learning approaches will be done through a graphical representation of the results. The goal and objective of this project is to predict the chances of diabetics then provide early treatment to patients, which will reduce the life-risk and cost of treatment. For this purpose a probability modeling and machine learning approach like Support Vector Machine algorithm Decision tree algorithm, Naive Bayes algorithm, Logistic regression algorithm are used to predict diabetics.

***Keywords*—** SVM (Support Vector Machine), Decision Tree, Naïve Bayes, Linear Regression, accuracy comparison, machine learning techniques, predicting data values, analysis and results**.**

I. INTRODUCTION

Diabetes Mellitus is a long-term illness characterized by the inability of the body to adequately metabolize glucose. The purpose of this study was to create an effective predictive model with high sensitivity and selectivity to better identify people at risk of Diabetes Mellitus using patient demographic data and laboratory results received during medical visits. The Diabetic is one of the main chronic diseases in the human body. It causes due to additional glucose levels found in the body. Due to lack of physical activity and less creation of insulin in the body. So that glucose level concentration is irregular. The patient glucose levels will be measured through proper diagnoses. Consequently patients need external insulin administration to control the blood glucose concentrations. Finally, based on the diagnoses reports doctors will plan treatment based on individual patient needs. If a patient needs external insulin doses it will be suggested by the doctor and can also be self-administered and recognize by the patients. According to the diagnosis reports analysis and Machine learning (ML) approaches doctor suggest appropriate treatment plan. Finally using ML predictive analytical methods diabetic disease reduction will be recommended by health care monitoring system respectively [1].

These chronic diseases play the most vital role in affecting the whole body. If a patient takes irregular food, unhealthy habits and no appropriate diagnosis it leads to serious health conditions such as loss of vision, heart attack, kidney failure etc. So it is very important to know about the patient health prediction from time to time. In the present scenario, the patient's urine and blood samples are collected and are diagnosed in the laboratory for the results. Or else using diabetic health wearable devices patient data can be collected and store the information in various health databases.

II. RELATED WORK

The goal and objective of this project is to predict and analyses diabetics work through machine learning algorithms and statistical methods. However, for the development of model researchers uses diabetic patient data. Patient data availability varies different dimensions and quality. Using the development methods and new technology, a wider amount of diabetic patient data is collected and examined constantly. With this requisite patient diabetes condition and early treatment is possible. Using these earlier prediction risks can be reduced in life.

III. Methodology

The proposed system design is applied through the supervised machine learning classification models. Using the following four models are implemented.

In Classification we are using the following four models.

1. Support Vector Machine (SVM).
2. Decision Tree.
3. Naïve Bayes.
4. Linear Regression

The Proposed system design model is detailed present in Fig: 1

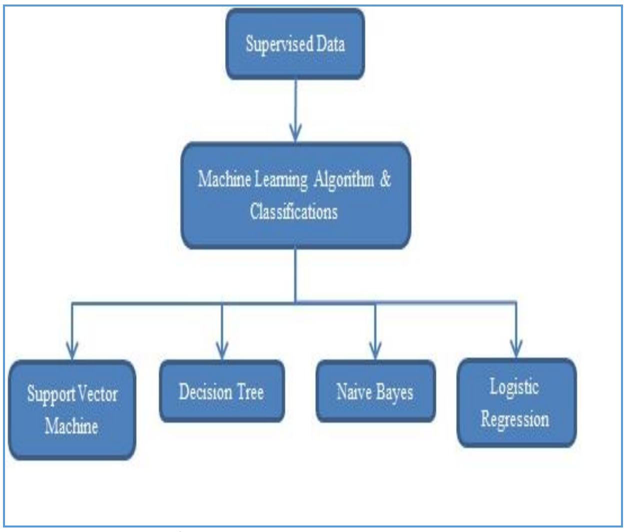


Fig 1 Flowchart for Supervised Machine Learning Algorithms used for Training Model

IV. Results and Discussion

We have identified many missing values and zero values in this dataset from Fig 7. Some of the rows contain zero for the columns preg, plas, insu and skin. A zero value is valid for the column preg but may not be valid for the columns like plas, insu and skin. Therefore, to deal with these missing values we have simply ignored the samples containing the missing values from fig.6 Replacing these missing values by the mean value is a legitimate approach to deal with missing values but, in this case it may bias the classification model which lead to more false positive rates.

As a result, the dataset's rows with missing values have been eliminated. There are 336 samples in the final dataset (after ignoring the rows containing missing values).Out of these 336 samples, 225 samples are of non-diabetic type and 111 are of diabetic type. This dataset is subjected to the four ML methods that were chosen. Before applying the algorithms, a 10-fold cross validation is utilized to partition the dataset into training and test sets. The results of the algorithms are shown in table 1.

As per the results, SVM and NB has achieved approximately same accuracy score of 77% from table 1. If execution time is considered, NB algorithm is faster than SVM algorithm with the default parameters. The other algorithms DT and LR have achieved accuracy scores 73% and 75% respectively which is less than SVM and NB algorithm.

The algorithms are executed with different parameter values but the default parameters produced better results as compared to other values.

V. Conclusion and Future Scope

So far analyzing the data, we discovered that Support Vector Machine, Decision Tree, Naïve Bayes and Logistic Regression are the best classifier. Machine Learning is used to forecast diabetic illness, and the model gives high-accuracy findings. When different classifiers are applied to the same data and the results are compared in terms of misclassification and correct classification rate, it is clear that Support Vector Machine is more accurate than Decision Tree, Naive Bayes algorithms and Logistic Regression,

We chose four popular classifiers based on their project performance. We choose one dataset from the Kaggle dataset repository to compare the performance of four learning algorithms in terms of classification for the current work.

As a part of future work, we are planning to use more efficient feature selection algorithms to select best relevant features from the dataset. Then ensemble approach can be used to improve the accuracy score without compromising the execution time.

VI. Prepare Your Paper Before Styling

Support Vector Machines Algorithm

Initially, in the earlier research study of support vector machines are used for the classification problems, but with only classification problems we can no longer meet people's needs. Due to the rapid development of computer technology, network technology, and database technology for the classification and management of large amounts of data. It's a hot research issue right now. There are two types of multi-class classification difficulties in existing treatment strategies [2].

To answer a multi-class classification problem, first construct a series of second-class classification problems in some fashion, then use the final class classification problems to solve the multi-class classification problem. This method is known as the multi-class SVM classification method.

The second method is to create a multi-class SVM from scratch, which may be used to address multi-class classification issues. There are two types of SVM.

SVM using linear approach: Linear SVMs are used for linearly indifferent data. That is, if one straight line can be used to classify a dataset into two classes, then the classifier is called a linear SVM classifier, with these data being linearly separable.

SVM using Non Linear approach: Non-linear SVMs are used for non-linearly indifferent data, but when it is not possible to classify datasets using straight lines, sorters that use these data as non-linear data are called non-linear SVM classifiers.

SVM Linear Approach:

An example can be used to explain how the SVM algorithm works. Let's imaginary we have a dataset with two tags (green and blue) and two qualities (x1 and x2). We need a classifier that can distinguish between green and blue coordinate pairs (x1, x2).Linear SVM approach is shown in the below Fig: 2.

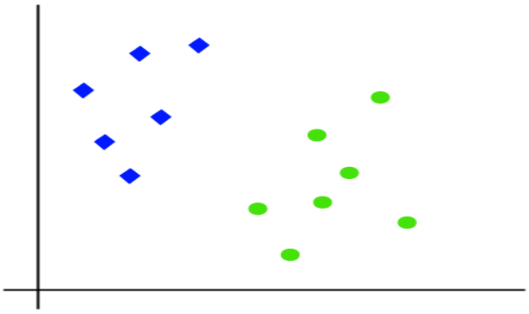


Figure: 2 Linear SVM

SVM Non Linear Approach:

We can separate data that is linearly structured using a straight line, but we cannot draw a single straight line for non-linear data. So we'll need to add another dimension to distinguish these data pieces. We've utilized two dimensions for linear data, x and y, so we'll add a third dimension, z, for non-linear data. It can be calculated using the formula z=x2 +y2.By tallying the third dimension, the sample space Non-Linear SVM approach is shown in the below Fig: 3.

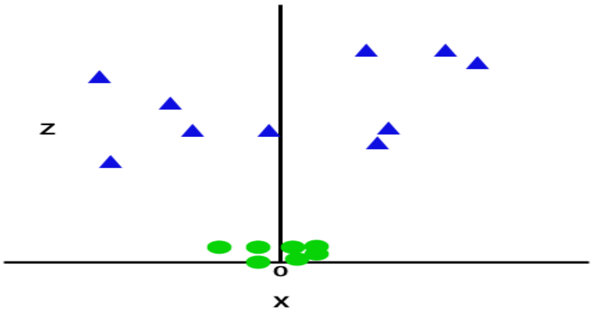


Figure: 3 Non-Linear SVM

SVM's Advantages and Disadvantages:

- When there is a distinct separation margin, it works best.
- It functions in three-dimensional spaces and when the number of dimensions exceeds the number of samples.
- It saves memory because the decision function only uses a subset of training points (called support vectors).
- It is connected to the SVC approach, which was created using the Python library.

Disadvantages

- It fails when we have a large dataset because the training time is greater.
- It also fails when the dataset has more noise, such as overlapping target classes.
- Instead of and not directly from the SVM, the probability estimates are produced using a costly five-fold cross-validation method.

Decision Tree Algorithm:

The supervised learning techniques include the Decision tree algorithm [3]. The results tree approach, unlike other supervised learning methods, may be utilized to tackle regression and classification problems. The result tree is used to build a training model that may be used to predict the class or value of target variables by learning conclusion rules learned from past data (training data).

In comparison to other classification algorithms, the result trees algorithm is easier to grasp. Each attribute corresponds to an inner tip of the tree, whereas each leaf tip corresponds to a class label.

The Sample Decision Tree Algorithm is shown in the below Fig: 4.

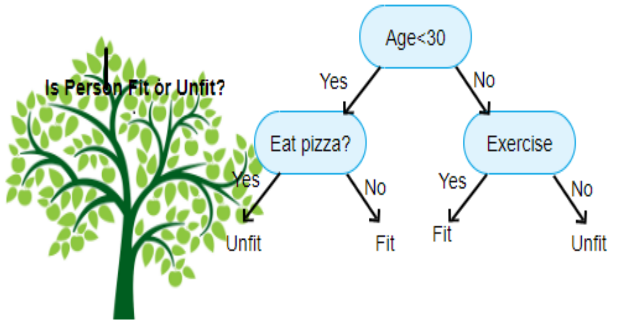


Figure: 4 Sample Model of Person fit or not using Decision Tree

Advantages:

- Decision trees are simple to understand. It yields a set of guidelines.
- It takes the same approach to decision-making that humans do in general.
- Visualizations can make it easier to understand a complex Decision Tree model. Even the most naïve individual can comprehend reasoning.

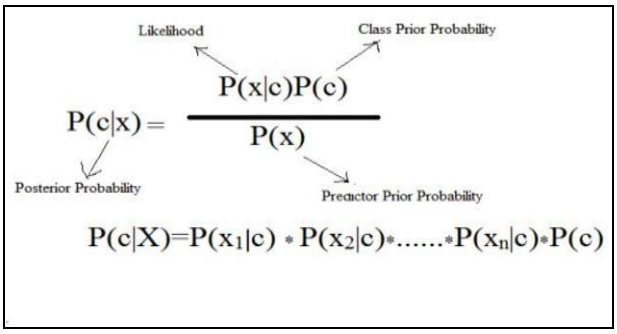
Disadvantages:

- In Decision Tree, there is a substantial risk of over fitting.
- When compared to other machine learning algorithms, it has a low prediction accuracy for a dataset.
- In a decision tree with categorical variables, information gain results in a biased response for qualities with more categories.

Naive Bayes Algorithm:

It's a classification technique based on the Bayes theorem and the predictor independence condition. In simple terms, a Naive Bayes classifier [4] asserts that the presence of one feature in a class has no bearing on the presence of any other feature. For Example: An apple is a fruit that is red, round, and about 3 inches in diameter. Even if these traits are conditional on one another or the presence of additional traits, they all contribute to the possibility that this fruit is an apple, which is why it is labelled as "Naive."

The Naive Bayes model is easy to build and works well with large data sets. Because of its simplicity, Naive Bayes is known to outperform even the most powerful classification algorithms.



$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$

$$P(c|X)=P(x_1|c) * P(x_2|c)*......*P(x_n|c)*P(c)$$

Equation: 1

Using P(c), P(x), and P(x/c), the Bayes theorem allows you to derive posterior probability P(c/x) from P(c), P(x), and P(x/c). Consider the following Naïve Bayes Algorithm is representation is shown in the above Equation: 1
As stated previously.

1. P(c/x) is the posterior probability of class (c, target) for a given predictor (x, attributes).
2. The class prior probability is P. (c).
3. The likelihood, P(x/c), is the chance that a predictor belongs to a specific class.
4. The prior probability of the predictor is P(x).

Advantages:

- Predicting the class of the test data set is simple and rapid. It's also good at predicting many classes.
- A Naive Bayes classifier outperforms conventional models like logistic regression when the assumption of independence is met, and it takes less training data.
- When contrasted to numerical input variables, it performs well with categorical input variables (s). The normal distribution is used for numerical variables.

Disadvantages:

- If a categorical variable in the test data set has a category that was not included in the training data set, the model will assign a probability of 0 (zero) and will not be able to predict.
- The probability outputs from predict probability, on the other hand, should be viewed with caution because naive Bayes is a reduced estimator.

Logistic Regression:

Logistic Regression is a classification algorithm. It is used to predict a binary outcome (1 / 0, Yes / No, True / False) given a set of independent variables. To express binary or categorical outcomes, dummy variables are utilized. Logistic regression is a term that can be used to describe a process. [5] As a type of linear regression in which the dependent variable is the log of odds and the outcome variable is categorical. In simple terms, it estimates the likelihood of an event occurring by fitting data to a log function.

The Logistic Regression Equation is derived as follows:
Generalized Linear Models are a bigger class of algorithms that includes logistic regression (glm). Nelder and Wedder burn created this model in 1972 as a way of applying linear regression to issues that were not immediately suitable for it. In reality, they suggested a variety of models (linear regression, ANOVA, Poisson Regression etc). As a special case, logistic regression was included [6].

The fundamental equation of generalized linear model representation is shown in the below Equation: 2

$$\boxed{\mathbf{g\ (E(y)) = \alpha + \beta x1 + \gamma x2}}$$

Equation: 2

Here, g () denotes the link function, E(y) is the target variable's expectation, and $\alpha + \beta x1 + \gamma x2$ denotes the linear predictor (,α, β, γ, to be predicted). The purpose of the link function is to 'connect' the linear predictor's expectation to the linear predictor's expectation.

Advantages:

- Easily adaptable to a variety of classes (multinomial regression).
- A probabilistic approach of class predictions that is natural.
- Easy to train - Extremely quick at categorizing unfamiliar records.

Disadvantages:

- Linear Decision Boundary.

Training Data: The experience gained by the algorithm is based on the observations in the training set. Each observation in a supervised learning problem consists of one or more observed input variables and an observed output variable. The enriched or labelled data you need to train your models is known as training data. To improve the accuracy of your model, you may just need to collect more of it. However, your data's prospects of being used are slim because, in order to construct a solid model, you'll need a lot of training data at scale [7].

Data Processing: Data pre-processing is a crucial stage in Machine Learning because the quality of data and the useful information that can be extracted from it has a direct impact on our model's ability to learn. It's crucial to pick the right parameters for your estimator. There are two elements to the training set: a training set and a validation set from Fig 6. The model can be trained based on the validation test results (for instance, changing parameters, classifiers).

Data Classification: Classification is a task that necessitates the application of machine learning algorithms to learn how to assign a class label to problem domain instances. Classifying emails as "spam" or "not spam" is an easy example. The results of classification predictive modelling algorithms are examined from Fig 7 Classification accuracy is a common metric for evaluating a model's performance based on projected class labels. Although classification accuracy isn't ideal, it's a solid place to start for a lot of classification problems [8].

Data Prediction: The technique of applying data analytics to create predictions based on data is known as predictive analytics. This method creates a predictive model for forecasting future events by combining data with analysis, statistics, and machine learning techniques. Predictive analytics uses these techniques to assess the likelihood of future outcomes based on historical data. Instead of only knowing what has happened, the goal is to make the greatest prediction of what will happen in the future [9].

Data Visualization: Data visualization is the process of converting information into a visual representation, such as a map or graph, in order to make data easier to comprehend and extract insights from. Data visualization's major purpose is to make it easier to spot patterns, trends, and outliers in massive data sets. By placing data in a visual context, such as maps or graphs, data visualizations helps us understand what it means. This makes the data more natural to understand for the human mind, making it easier to see trends, patterns, and outliers in vast data sets. The comprehensive study of System Architecture for Diabetics Prediction is shown below in Fig: 5.

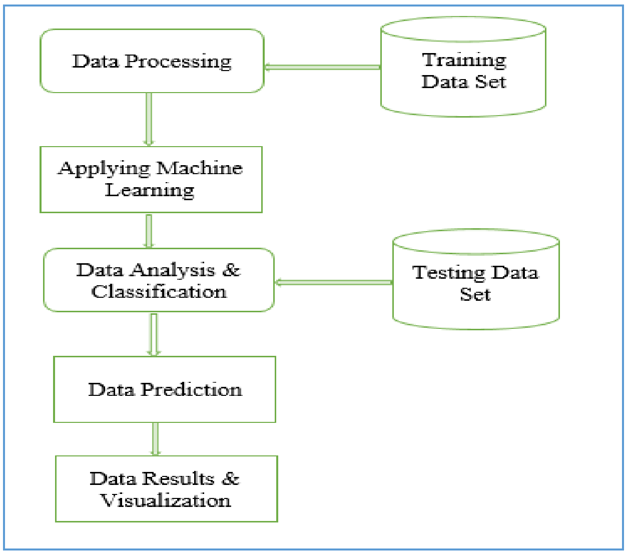
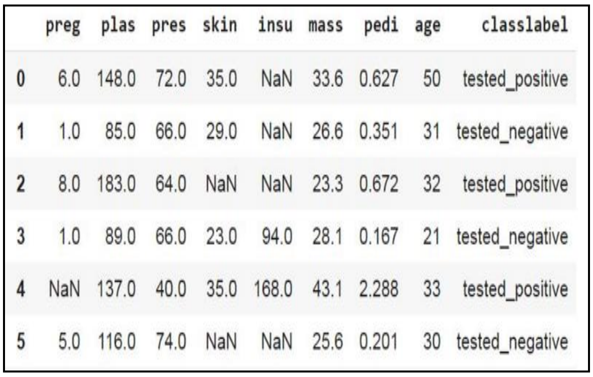

Data Processing
Training Data Set
Applying Machine Learning
Data Analysis & Classification
Testing Data Set
Data Prediction
Data Results & Visualization

Fig 5: System Architecture for Diabetics Prediction.

Data Processing Steps:
The steps for pre-processing are as follows:
1. Dealing with Null Values
2. Standardization is a term that refers to the process of bringing
3. Taking Care of Categorical Variables
4. Encoding in a Single Step
5. Multi linearity is a term that refers to the fact that there are multiple.

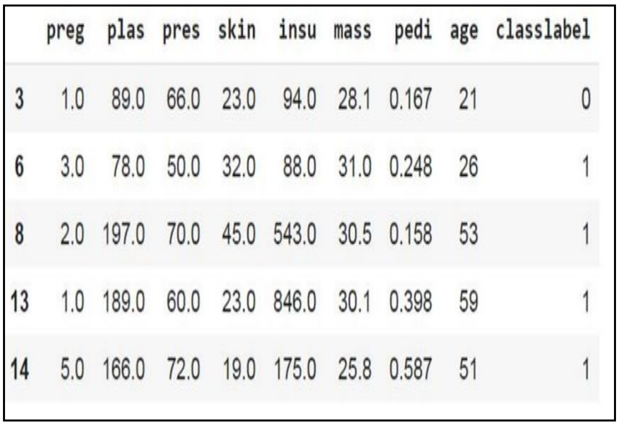
The Sample data set is collected from Kaggle Data repository for system designing [10]. The collected data acquired with missing values is shown in Fig: 6.



	preg	plas	pres	skin	insu	mass	pedi	age	classlabel
0	6.0	148.0	72.0	35.0	NaN	33.6	0.627	50	tested_positive
1	1.0	85.0	66.0	29.0	NaN	26.6	0.351	31	tested_negative
2	8.0	183.0	64.0	NaN	NaN	23.3	0.672	32	tested_positive
3	1.0	89.0	66.0	23.0	94.0	28.1	0.167	21	tested_negative
4	NaN	137.0	40.0	35.0	168.0	43.1	2.288	33	tested_positive
5	5.0	116.0	74.0	NaN	NaN	25.6	0.201	30	tested_negative

Fig 6: Sample Data Set with Missing Values

The dataset's total number of samples is 768 which contains missing values for some of the attributes. In order to make the dataset meaningful, we have dropped the missing values from the dataset which results in 336 number of samples. There are so many other methods available to deal with missing values, but we have ignored the missing values in order to minimize the training time. The dataset without missing values is shown Fig: 7.

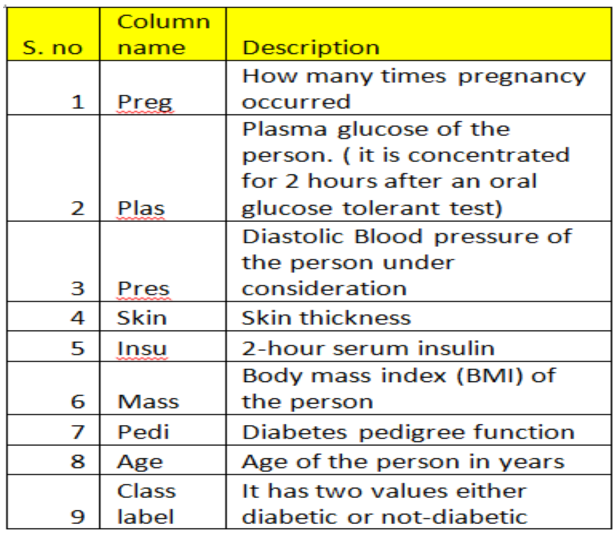


	preg	plas	pres	skin	insu	mass	pedi	age	classlabel
3	1.0	89.0	66.0	23.0	94.0	28.1	0.167	21	0
6	3.0	78.0	50.0	32.0	88.0	31.0	0.248	26	1
8	2.0	197.0	70.0	45.0	543.0	30.5	0.158	53	1
13	1.0	189.0	60.0	23.0	846.0	30.1	0.398	59	1
14	5.0	166.0	72.0	19.0	175.0	25.8	0.587	51	1

Fig 7: Sample Data Set without Missing Values

Data Set Descriptions
The dataset used in this work contains 768 rows and 9 columns. This diabetic dataset is collected from the online sources. Pre-processing was done using Python programming language.

The columns with description of the dataset are shown in below Fig: 8.



S. no	Column name	Description
1	Preg	How many times pregnancy occurred
2	Plas	Plasma glucose of the person. (it is concentrated for 2 hours after an oral glucose tolerant test)
3	Pres	Diastolic Blood pressure of the person under consideration
4	Skin	Skin thickness
5	Insu	2-hour serum insulin
6	Mass	Body mass index (BMI) of the person
7	Pedi	Diabetes pedigree function
8	Age	Age of the person in years
9	Class label	It has two values either diabetic or not-diabetic

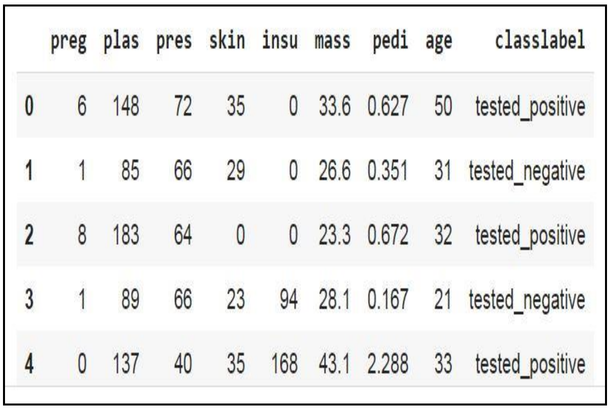
Fig 8: Data Set Descriptions

While pregnant, Glucose level: In an oral glucose tolerance test, plasma glucose concentration over 2 hours. Diastolic blood pressure is the highest level of blood pressure (mm Hg), Triceps skinfold thickness is measured in millimeters (mm), BMI: Body mass index (weight in kg/ (height in m) 2), Insulin level: 2Hour serum insulin (mu U/ml), Diabetes Pedigree's Purpose: The pedigree function of diabetes (a function which scores the likelihood of diabetes based on family history, Age is a factor (years). The Diabetes Dataset's description and ranges are provided in the data set [11].

1. Coding And Implementation

The coding or programming phase's purpose is to convert the system design created during the design phase into code in a programming language that can be run by a computer and performs the calculation defined by the design.

Both testing and maintenance are affected by the coding process. The purpose of coding should not be to lower the cost of implementation, but to lower the cost of subsequent phases. In other words, the purpose is not to make a programmer's job easier. Rather, the goal should be to make the testers and maintainer's jobs easier. Using class labeling design the data set information can be easily verifiable as shown in the Fig: 9.



```
   preg  plas  pres  skin  insu  mass   pedi  age        classlabel
0     6   148    72    35     0  33.6  0.627   50   tested_positive
1     1    85    66    29     0  26.6  0.351   31   tested_negative
2     8   183    64     0     0  23.3  0.672   32   tested_positive
3     1    89    66    23    94  28.1  0.167   21   tested_negative
4     0   137    40    35   168  43.1  2.288   33   tested_positive
```

Fig 9: Check the information of the Data Set.

Total class count of the diabetic data effected with diabetic test positive and diabetic test negative set can be easily identified in the Fig: 10.

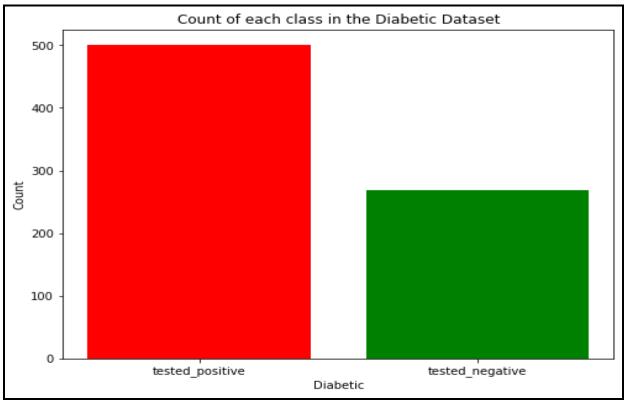


Fig 10: Count of each Diabetic Data Set

Accuracy Comparison:
When they are different classifiers applied together to the same data and the results are compared in terms of misclassification and correct classification rate, it is clear that Support Vector Machine is more accurate then Logistic Regression, Decision Tree, and Naive Bayes with general computational comparison of algorithm results are shown in the Fig: 11.

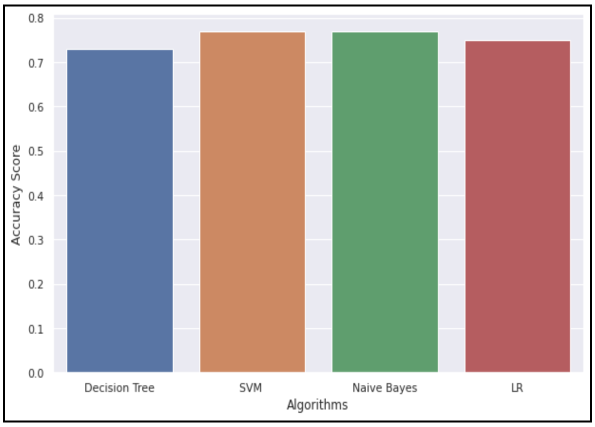
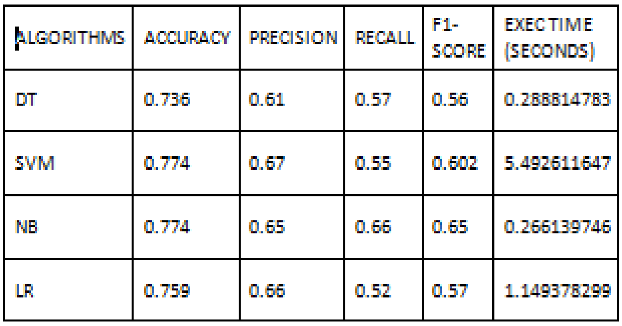


Fig 11: Accuracy Comparison Graph

Result Table:

You can find the results from the below Table: 1.

Table 1: Accurate outcomes of algorithm comparison



ALGORITHMS	ACCURACY	PRECISION	RECALL	F1-SCORE	EXEC TIME (SECONDS)
DT	0.736	0.61	0.57	0.56	0.288814783
SVM	0.774	0.67	0.55	0.602	5.492611647
NB	0.774	0.65	0.66	0.65	0.266139746
LR	0.759	0.66	0.52	0.57	1.149378299

Decision Tree outcomes are displayed in Fig: 12.

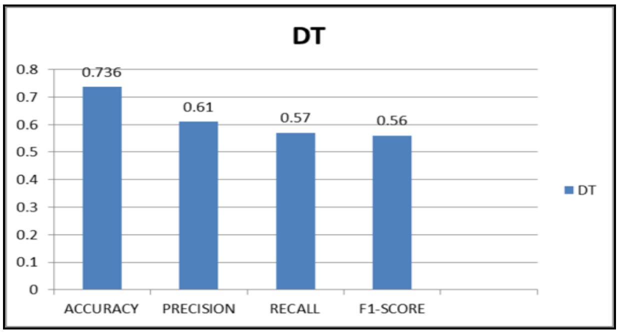


Fig 12: Decision Tree Algorithm Results

SVM Results outcomes are displayed in Fig: 13.

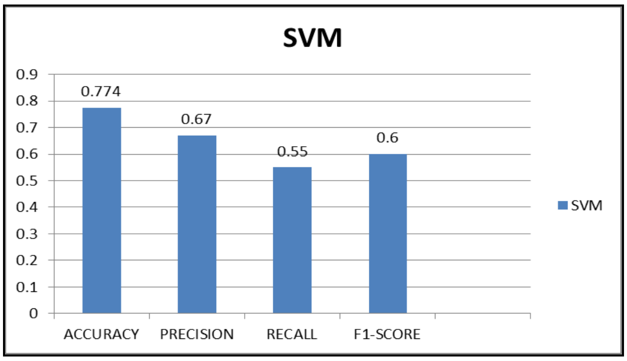


Figure 13: SVM Algorithm Results

NB Results outcomes are displayed in Fig: 14.

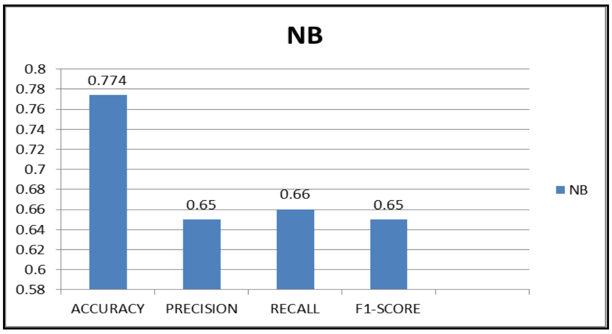


Fig 14: NB Algorithm Results

Logistic Regression Results outcomes are displayed in Fig: 15.

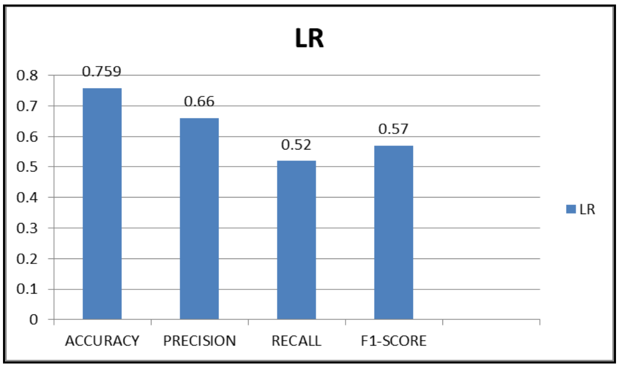


Figure 15: Linear Regression Algorithm Results.

Execution time comparison of all algorithms are displayed in Fig: 16.

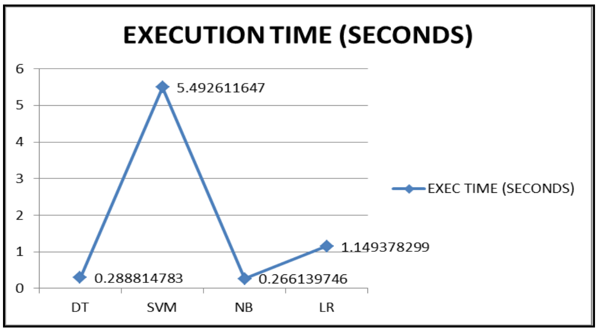


Figure 16: Comparison of all algorithms

ACKNOWLEDGMENT

The preferred spelling of the word "acknowledgment" in America is without an "e" after the "g". Avoid the stilted expression, "One of us (R. B. G.) thanks . . ." Instead, try "R. B. G. thanks". Put sponsor acknowledgments in the unnum-bered footnote on the first page.

REFERENCES

[1] S.Deepti, S.Dilip Singh, "*Prediction of Diabetes using Classification Algorithms",* Procedia Computer Science, Vol.**132,** pp. 1578-1585, **2018**.

[2] Larabi-Marie-Sainte, S, Aburahmah, L, Almohaini, R, & Saba, T, "*Current techniques for diabetes prediction* review and case study", *Applied Sciences*, Vol.**9**, Issue.**21**, pp.460, **2019**.

[3] Rodríguez-Rodríguez, I Rodríguez, J.V.Woo, W. L, Wei, B, & Pardo-Quiles, D. J, "*A Comparison of Feature Selection and Forecasting Machine Learning Algorithms for Predicting Glycaemia in Type 1 Diabetes Mellitus*", *Applied Sciences*, Vol.**11**, Issue.**4**, pp.1742, **2021**.

[4] Nedyalkova, M., Madurga, S., & Simeonov, V. Combinatorial "*k-means clustering as a machine learning tool applied to diabetes mellitus type 2*", International Journal of Environmental Research and Public Health, Vol.**18**, *Issue* **4**, pp.1919, **2021**.

[5] Daniel, P. "*An application of the free moment for the diabetic patients' classification-a pilot study",* E-Health and Bioengineering Conference (EHB), pp. 1-4, IEEE, **2015**.

[6] Taser, P.Y, "*Application of Bagging and Boosting Approaches Using Decision Tree-Based Algorithms in Diabetes Risk Prediction*", Multidisciplinary Digital Publishing Institute Proceedings Vol. **74**, No. **1**, p. 6, **2021**.

[7] Kanchan, B. D, & Kishor, M.M, "*Study of machine learning algorithms for special disease prediction using principal of component analysis*", International conference on global trends in signal processing, information computing and communication (ICGTSPICC), pp. 5-10, IEEE **2016**.

[8] Mohanty, K. K, Barik, P. K, Barik, R. C, & Bhuyan, K. C, "*An efficient prediction of diabetic from retinopathy using machine learning and signal processing approach*", International Conference on Information Technology (ICIT) pp. 103-108, IEEE, **2019**.

[9] Shi, G, Zou, S, & Huang, A, "*Glucose-tracking: A postprandial glucose prediction system for diabetic self-management*", 2nd International Symposium on Future Information and Communication Technologies for Ubiquitous HealthCare Ubi-HealthTech pp. 1-9, IEEE **2015**.

[10] www.kaggle.com/uciml/pima-indians-diabetes-database.

[11] Godi, B, Viswanadham, S, Muttipati A. S, Samantray O. P, & Gadiraju S. R, *"E-healthcare monitoring system using IoT with*

machine learning approaches", International Conference on Computer Science, Engineering and Applications (ICCSEA), pp. 1-5, IEEE **2020**.

AUTHORS PROFILE

Ms.Ommi Ramu has completed her Bachelor of Engineering in the branch of Computer Science Engineering from Raghu Institute of technology affiliated by Jawaharlal Nehru Technological University, Kakinada. She is now pursuing her Masters of Engineering (CSE) in Raghu Institute of Technology, Visakhapatnam, and Andhra Pradesh, India. Her interests includes Machine Learning, Artificial Intelligence, Python, Java, and Cyber Security. She has done her masters project using Machine learning approaches.

Mr. Brahmaji Godi has completed MSc.IS from Andhra University in 2003, M.Tech CST from Andhra University in 2009.Currently working as Assistant Professor in Department of Computer Science and Engineering at Raghu Institute of Technology Visakhapatnam. He has published 5 research papers from reputed international journals which are approved by UGC, Scopus index and IEEE. His main research work focuses on Internet of Things, Artificial Intelligence, Computer Vision, Human Computer Interaction, Machine Learning and Wireless networks. He has 13 years of teaching experience. He is having member ship in International Association of Engineers, The Society of Digital Information and Wireless Communication.

Mr.Om Prakash Samantray has completed his B.Tech in Information Technology from Biju Patnaik University of Technology, Odisha in 2006 and completed M.Tech in Computer Science and Engineering from BPUT, Odisha in 2010. He has submitted his Ph.D. Dissertation at Berhampur University, Odisha in Dec, 2020. He has 12 years of teaching experience in Computer Science and Engineering and presently working as Assistant Professor with Raghu Institute of Technology, Visakhapatnam. He has published 14 research articles in international conferences and journals. His research interests include; Data Mining, Machine Learning, Computer Vision and Network Security.