Research Journal of Multidisciplinary Engineering Technologies
(An Open access International peer reviewed referred journal)
ISSN: 2945-4158

Research Article

Diabetes Prediction Using Logistic Regression Machine Learning Algorithm

Ramesh Prasad Bhatta[1], Dipendra Kumar Air [2]
[1,2] Assistant Professor , Central Department of CSIT, Far Western University, Nepal

Abstract: Diabetes is a serious worldwide health issue that is becoming more of a problem in Nepal because of its high risk of death and other complications. This study develops an early prediction model using logistic regression, a widely applied machine learning classification technique in clinical research. The model was implemented in Python IDE with data from the Pima Indians Diabetes Database, which includes 768 patient records comprising eight independent features and one outcome variable. Exploratory data analysis was performed to extract insights and visualize trends in the dataset. To address class imbalance, the Synthetic Minority Oversampling Technique (SMOTE) was applied, generating synthetic samples for the minority class. Model evaluation using a confusion matrix demonstrated satisfactory results, achieving an accuracy of 77%, precision of 75%, recall of 77%, and an F1-score of 76%. To further enhance performance, hyperparameter tuning was conducted using the grid search method. The model after grid search improved outcomes, reaching an accuracy of 82%. These findings suggest that logistic regression, supported by data preprocessing, resampling techniques, and hyperparameter optimization, can serve as an effective tool for early detection of diabetes, thereby supporting timely intervention and improved healthcare outcomes.

Keywords: Diabetes, Machine Learning, Prediction, Regression, Accuracy

How to cite this article: Bhatta RP, Air DK . Diabetes Prediction Using Logistic Regression Machine Learning Algorithm. **Research Journal of Multidisciplinary Engineering Technologies.2025 Dec; 4(6):1-14**

Source of support: Nil.

Conflict of interest: None

DOI: doi.org/10.58924/rjmet.v4.iss6.p1

Received: **12-12-2025**
Revised: **15-12-2025**
Accepted: **22-12-2025**
Published: **29-12-2025**

1. Introduction

Diabetes is considered as major healthcare issue that is affecting the world at a rapid and alarming rate [1]. "Diabetes mellitus is deadliest and is caused by a set of metabolic disorders that occur when the body cannot produce any or enough insulin or cannot effectively use the insulin it produces" [2]. When a person has abnormally high blood glucose levels because of either inadequate insulin manufacturing or an inappropriate cell response to insulin, they have diabetes mellitus, one of the metabolic illnesses.

Millions of people worldwide have been afflicted with diabetes mellitus, a diverse and non-communicable disease, at an alarming rate.

Patients with diabetes and those at risk for developing the disease often exhibit the following early symptoms and indicators: increased thirst, weariness, weight gain, lightheadedness, discolored skin, and sexual deficiency. Diabetes mellitus can adversely affect the human organs like brain, nerve damage, heart, kidney, eyes, skin so on and so forth. Diabetes is causing a patient's pancreatic beta cells to be pathologically destroyed. [3].

Types of Diabetes

1. **Type 1-Diabetes mellitus (T1DM):** Another name for it is diabetes that is not insulin-

dependent. It is the most common kind of diabetes and is identified by the body's inadequate production of insulin. The illness can strike at any age, although children and teenagers are the ones who get it most often [4].In this type of diabetes, pancreas will produce insulin that helps human organs to energize through sugar level in the blood cells. But there may be a chance that pancreas might be producing little amount of insulin or no insulin. Insulin injections are commonly used for controlling Type-1 diabetes. Type-1 diabetes is common in any aged people but it most affects the people among under age 30. Particularly if the patient's heredity having Type-1 diabetes will lead to higher risk. Statistically below 10% of the people impacted by this particular form of diabetes.

2. **Type 2-Diabetes mellitus (T2DM)**: It is the most prevalent kind of diabetes and is distinguished by the body's insufficient synthesis of insulin. All age groups are affected, and patients frequently show signs of obesity, overweight, urination, etc., which are associated with the Insulin resistance [5].Historically, the adults are most commonly affected by Type 2 diabetes. Statistics revealed that betwixt 90-95 percent of the people will be affected by type 2 diabetes. Diet through weight management and exercise are the common ways to control Type 2 diabetes. Yet, medications or injections may be considered as remedy for lowering the glucose level

3. **Gestational diabetes mellitus (GDM):** the kind of diabetes that causes pregnant women to have hyperglycemia. This kind of diabetes raises the mother's and the fetus's risk of developing type-2 diabetes. [6].In general, the pregnant women will have Gestational diabetes who never had a diabetes in their lifetime. The glucose level will be high when the women get pregnancy. The baby has higher glucose level at the time of pregnancy. Changes in hormone will also leads to high glucose level in blood that affects the action of insulin.

4. **Prediabetes (PD):** Genetic abnormalities that result in increased insulin production are the cause of this form of diabetes., side effects of chemicals, or increase in other hormonal levels in the body.A good lifestyle is likely to be most beneficial in preventing the progression of T2DM at early stages before the medication and other treatments. Lifestyle habits and demographic factors are examined and reported the main indicators that play an important role to control and manage Type-2 Diabetes Mellitus.Diet and exercise play an important role to avoid or manage the T2DM, it can reduce the complications of even those people who are at high risk of being involved towards disease

The diets and exercise are crucial in maintaining, preventing, and controlling diabetes, particularly Type-2 Diabetes Mellitus. The material now in publication shows how diet and exercise regimens relate to many chronic illnesses. Figure illustrates the several elements that influence diabetes.

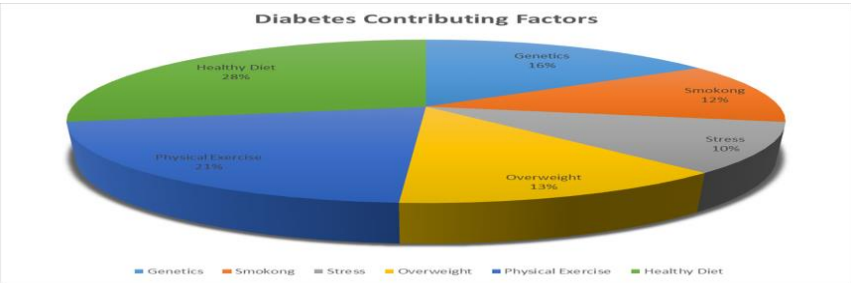


Fig.1: Factors Contributing Diabetes

Diabetes is now among the most prevalent illness, presently faced by humans, all over the world (more than 425 million in 2017). In India, around 82 million individuals were

recorded as diabetic in 2017. 90% of diabetic patients suffer from T2DM. The main cause of T2DM is unhealthy diet habits with least or no exercise plans that also lead to obesity, overweight, high blood pressure, and finally diabetes. Diabetes is not just a disease; it is an ailment that drastically degrades the life expectancy and living standards of a human being. An early diagnosis and proper lifestyle, in terms of diet and exercise, can be the key to minimize its effect. It can also help to reduce the rate of increase in the number of diabetic patients (expected to rise by 48% i.e., 629 million in 2045).Millions of individuals worldwide have been impacted by diabetes, a chronic and deadly illness. The number of people with diabetes has dramatically increased during the last ten years, making the condition a global concern. [7].

463 million persons between the ages of 20 and 79 (9.3% of the global population) currently have diabetesbased on a study issued by the International Diabetes Federation (IDF) Atlas 2019. Additionally, it is predicted to impact 700 million people by 2045 and 578 million by 2030. According to estimates, diabetes mellitus caused 4.2 million deaths globally in 2019 [8]. According to the IDF Atlas 2019, India is predicted to have the highest number of diabetics in the world by 2045, with [134.3-165.2] million [8]. Figure 1.3 lists the ten countries with the highest rates of diabetes. Insulin generally has a considerable effect on blood glucose levels.

Statistics of Diabetes

[9] India is known as Diabetes capital of world and the statistics reveals that 74.9 million people is affected by diabetes by 2021 and it will raise to 124.9 million by 2045 in India. Hence, India is facing challenge to overcome the scenario. But still, the medical analysts advised that early prediction and right decision helps the diabetes patients to overcome the diabetes affections and also they can lead an ordinary life. World Health Organization (WHO) disclosed that 3.5 million demises were happened in India because of high blood sugar. WHO, reveals that demise percentage will be 90 percent between 2017 to 2030. Also, it is predicted that type-2 disease will rise to 438 million in 2030 whereas it is estimated that in India, 58 percent will rise up to 87 percent in 2030. In general, insulin strongly controls the blood glucose level. In the Southeast Asia region, out of 88 million people with diabetes, India contributes 77 million people, which is expected to increase to 134.2 million in 2045, as shown in figure above [2]. According to the International Diabetes Federation (IDF), diabetes prevalence among adults in Nepal reached 7.7% in 2024, affecting approximately 1.3 million individuals. The burden has increased sharply, rising from about 488,200 cases in 2011 to 1.1 million in 2021, with projections estimating 2.4 million cases by 2050 [6]. Major risk factors include obesity, hypertension, elevated triglyceride levels, older age, male sex, and urban lifestyle. Globally, an estimated 463 million people were living with diabetes mellitus (DM) in 2019, marking a 62% increase since 2009. This figure is expected to grow by 25% by 2030 and 51% by 2045.

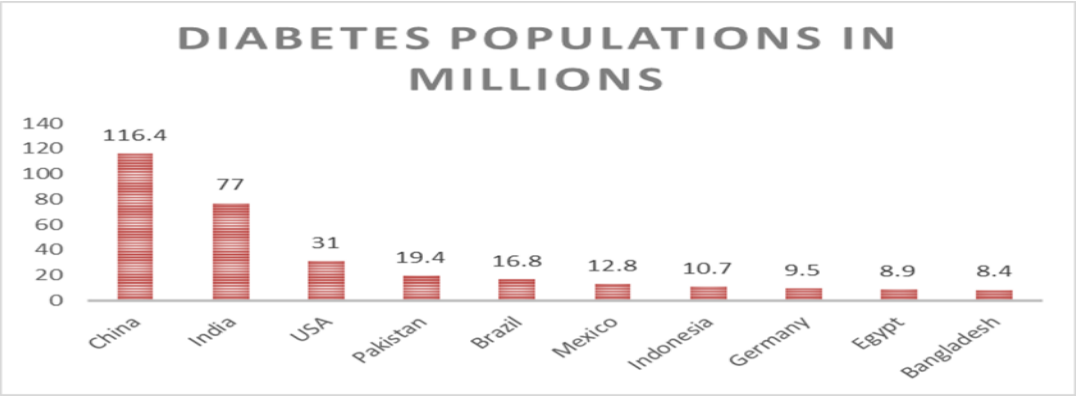


Figure: 1.4 Top 10 countries with Diabetes People

2. Machine Learning in Diabetes Prediction

Machine learning (ML) concepts originated from pattern recognition have the ability to discover without programming to perform specific tasks. Samuel invented "ML" in the early1950s [4]. In ML, 'A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E.' [5]. Large, complex data sets can be analyzed using ML to find patterns and create hypotheses by learning from previous examples [6]. The intention of using AI and ML in healthcare is to increase the diagnostic accuracy and effectiveness of therapy and help clinicians in their practice of patient management with improved outcomes

In machine learning, data is crucial. The ever-increasing volume of data, the cost of data storage, the variety of data types, the complexity of data, the heterogeneity of data, and the utilization of multiple sources are just a few of the challenges that impede the vital process of data collection. Even though the field of machine learning has accomplished a great deal, human labor cannot be entirely replaced by machines.The application of machine learning techniques in the medical field is growing. To predict diabetes, numerous researchers have employed a variety of machine learning and deep learning techniques and algorithms.

3. Methodology

The prediction model intended to ascertain the probability of DM occurrence is presented in this section. Model building and diabetes prediction were conducted using the Pima Indian Diabetes Dataset (PIDD), which was obtained from the UCI website.The working of this study was illustrated in the following figure 2.A logistic regression algorithm was used to predict diabetes.

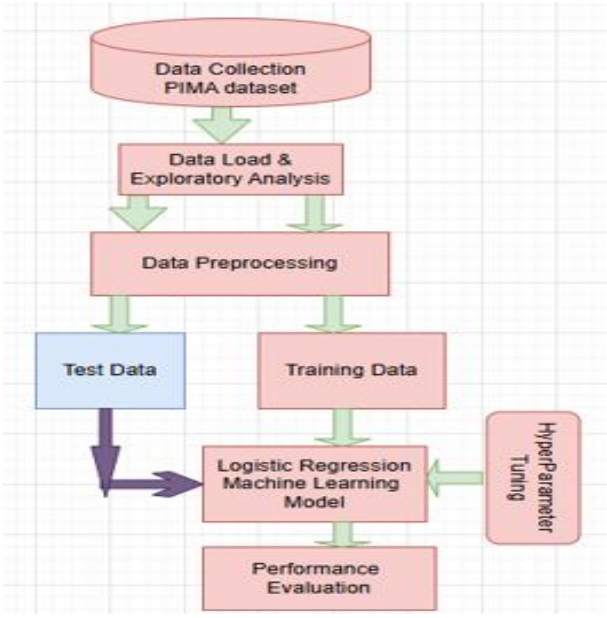


Figure 1: Working of Proposed System

3.1 Dataset Description

The dataset was taken from Pima Indians Diabetes Database. The dataset consists of eight independent variables and one dependent variable (outcome variable) as shown in Table 1.

Table 1: Features used in PIMA Diabetes dataset

Attributes	**Inference**
Insulins	Amount of insulin in a 2 h serum test
Number of pregnancies	Number of times the person has been pregnant
Glucose concentration	Blood glucose level on testing
Age	Age of the person
BMI	Body mass index
Skin thickness	Skin fold thickness of the triceps
Diastolic blood pressure	Diastolic blood pressure
Diabetic pedigree function	function that assesses the likelihood of diabetes based on family history
Outcome	The person is predicted to have diabetes or not

3.2 **Data Exploration**

Data exploration involves getting insights about the data and finding the correlation between the features. Dataset consists of 268 patients with diabetes and the remaining 500 patients without diabetes. The heatmap displayed in Figure 3, shows the correlation between the features of Dataset. The lighter colors represent more correlation and the darker colors represent less correlation

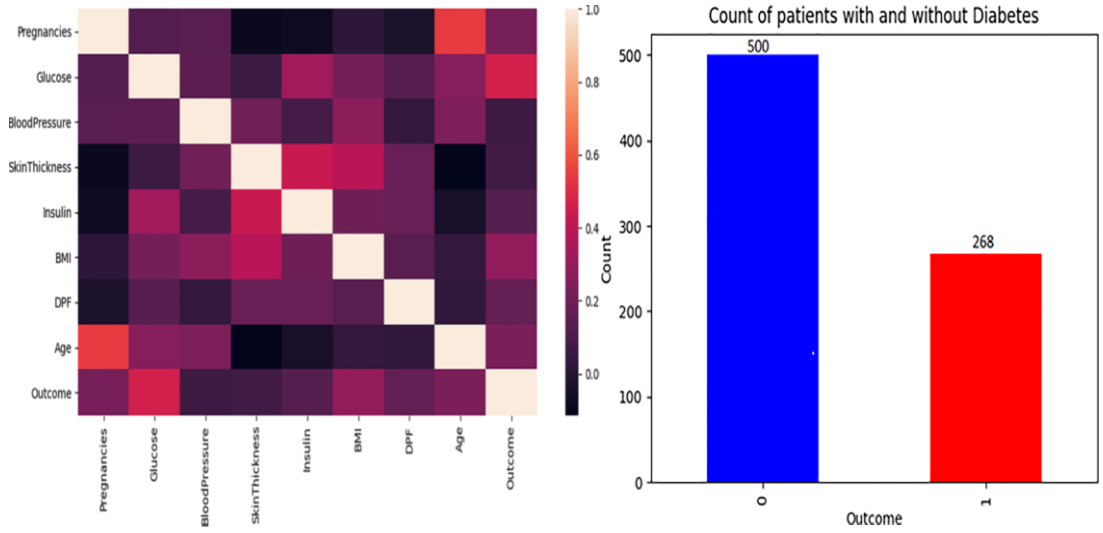


Fig.3: Correlation between dataset features

The Fig. 3 shows a bar plot displaying count of patients with and without Diabetes.

3.3 Data Preprocessing

The process of converting unstructured data into a format appropriate for machine learning model training is known as data preprocessing in machine learning. Data preprocessing handles issues like missing values, inconsistencies, noise, and irrelevant features

Handling missing values:
The missing data values were replaced with the median value of each variable so that each data in the dataset variable had an absolute value. The missing values, null values and values equal to zero for the predictor variables need to be identified in the dataset. The predictorvariables/features cannot have a zero value except for certain features, like for example the 'Pregnancies' feature in Dataset. These valuesneed to be replaced with the mean values of the column.

Imbalanced Dataset: The number of diabetic patients is 268 people (34.9%), while the number of non-diabetes patients is 500 (65.1%). SMOTE technique was used to overcome the imbalanced data. SMOTE is an oversampling technique in which a synthetic sample is generated for the minority class to help overcome the overfitting problem found in random oversampling.

3.4 **Model Development Using Logistic Regression Algorithm**

Logistic regression models the relationship between categorical and covariate response variables. Logistic regression is a linear model that is more suitable for problems classification.In the literature as logit is used as regression., maximum-entropy classification (MaxEnt). In logit, the probabilities describing the possible outcomes of a single experiment are modeled using the logistic function. Logistic regression models can be binary, one-vs-rest, or multinomial logistic regression with l1, l2 or elastic-net regulation [27]. Binary logistic regression estimates the probability of the availability of a binary variable characteristic, given the covariate value. For example, Y is a binary response variable with Yi = 1 if the character is available, and Yi = 0 if the character is unavailable and data [$Y1$, $Y2$, Yn] are independent. The value of πi can be used to be a successful probability of a logistic regression. In addition, x = (x_1, x_2, x_p) value is also considered a set of variables that can be discrete, continuous, or a combination of both. The π_i logistic function is given by (1) and (2).

$$\text{logit}(\pi_i)=\log\left(\frac{\pi i}{1-\pi i}\right)=(\beta_0+\beta_1 x_{i1}+\beta 2 x_{i2}+\ldots+\beta_p x_{ip} \quad \ldots\ldots\ldots\ldots \quad (1)$$

$$\pi i = \frac{\exp(\beta_0+\beta_1 x_{i1}+\beta_2 x i_2+\ldots+\beta_p x_{ip})}{1+\exp(\beta_0+\beta_1 x_{i1}+\beta_2 x i_2+\ldots+\beta_p x_{ip})} \quad \ldots\ldots\ldots \quad (2)$$

$$=\frac{\exp(x_i'\beta)}{1+\exp(x_i'\beta)}=\Lambda(x_i'\beta)$$

In this equation, π_i represents the probability that the sample falls under a particular category of the dichotomous response variable, commonly referred to as the probability of success. It is obvious that $0 \le \pi_i \le 1$. Λ (.) is a logistic cumulative distribution function (CDF) with

$(z)=e^z/(1+e^z)$
$=1/(1+e^z)$

and β^s represents parameters' vector to be estimated. The equation $\pi_i/1-\pi_i$ is called the odds ratio or relative risk [28]. This study used a binary logistic regression algorithm since the output was a value of 0 and 1, which was used to detect whether a person has diabetes. An output value of 0 suggests that a person does not have diabetes and an output of 1 indicates that the person has diabetes.

3.5Training and testing:

The data used in this study is divided 70:30, meaning that 70% is used for training and 30% is used for testing. The predictions are made and their correctness is verified using a logistic regression technique.

3.6 Model Evaluation

A confusion matrix is a matrix that summarizes the performance of a machine learning model on a set of test data. It is often used to measure the performance of classification models, which aim to predict a categorical label for each input instance. True Positive (TP): Correctly predicted positive cases. False Positive (FP): Incorrectly predicted positive cases (actually negative). True Negative (TN): Correctly predicted negative cases. False Negative (FN): Incorrectly predicted negative cases (actually positive

Accuracy (ACC):

Accuracy is computed as the number of all correct predictions divided by the total number of the dataset, which is the number of patients that are identified correctly in total in our case.

$$\text{Accuracy} = \frac{(TP+TN)}{(TN+FP+FN)} \quad \ldots\ldots. \quad 3$$

Precision:

Precision is computed as the number of correct positive predictions divided by the total number of positive predictions.

$$\text{Precision} = TP/(TP+FP) \quad \ldots\ldots. \quad 4$$

Recall(Sensitivity)

Recall is computed as the number of correct positive predictions divided by the total number of positives. it is also called Sensitivity or true positive rate (TPR).

$$\text{Recall} = TP/(TP+FN) \quad \ldots\ldots. \quad 5$$

F1 score:

The F1-score is the harmonic mean of precision and recall, providing a balance between the two metrics. It is calculated as twice the product of precision and recall divided by the sum of precision and recall. it provides the quality of prediction.

$$\text{F1 score} = 2*(\text{Precision}*\text{Recall})/(\text{Precision} + \text{Recall}) \quad \ldots\ldots. \quad 6$$

3.7 Hyperparameter Tuning

To improve model performance, hyperparameter tuning selects the optimal set of hyperparameters. The values of the model arguments that comprise hyperparameters are set prior to the start of the learning process. In this study, a grid search approach that aggregated input values in hyperparameters was used. The grid search method would search for every possible combination and then choose the best one based on the highest cross-validation value. The C value and the penalty were the two hyperparameters used in this investigation. The penalty applied the l1 and l2 regulations, and the C value was the inverse of the regularization strength (l2 was the default value).

4. Results and Discussion

Load and Read the Data

Python was used to load datasets in.csv format into the Jupiter system. 768 patient records were found, all of whom were female and at least 21 years old. There are nine variables in the dataset. It includes one dependent variable, called outcome, and eight independent variables, including age, diabetes, blood pressure, glucose, skin thickness, insulin, BMI, diabetes pedigree function, and DPF. data.head() was used to inspect the dataset, and the results revealed that a number of variables had values of 0, indicating a missing value.

Pregnancies	Glucose	BP	Sthick	Insulin	BMI	DPF	Age	Outcome
0	6	148	72	35	0	33.6	0.627	50	1
1	1	85	66	29	0	26.6	0.351	31	0
2	8	183	64	0	0	23.3	0.672	32	1
3	1	89	66	23	94	28.1	0.167	21	0
4	0	137	40	35	168	43.1	2.288	33	1
..	...	...	..	...	...	...	...	...	...
763	10	101	76	48	180	32.9	0.171	63	0
764	2	122	70	27	0	36.8	0.340	27	0
765	5	121	72	23	112	26.2	0.245	30	0
766	1	126	60	0	0	30.1	0.349	47	1
767	1	93	70	31	0	30.4	0.315	23	0

[768 rows x 9 columns]

Additionally, the image demonstrates that a number of variables have empty data values. The top five data points indicate that the variables for insulin and pregnancies have null values (0). To make the data modification process easier, the median value of each variable would be used to replace the empty data values

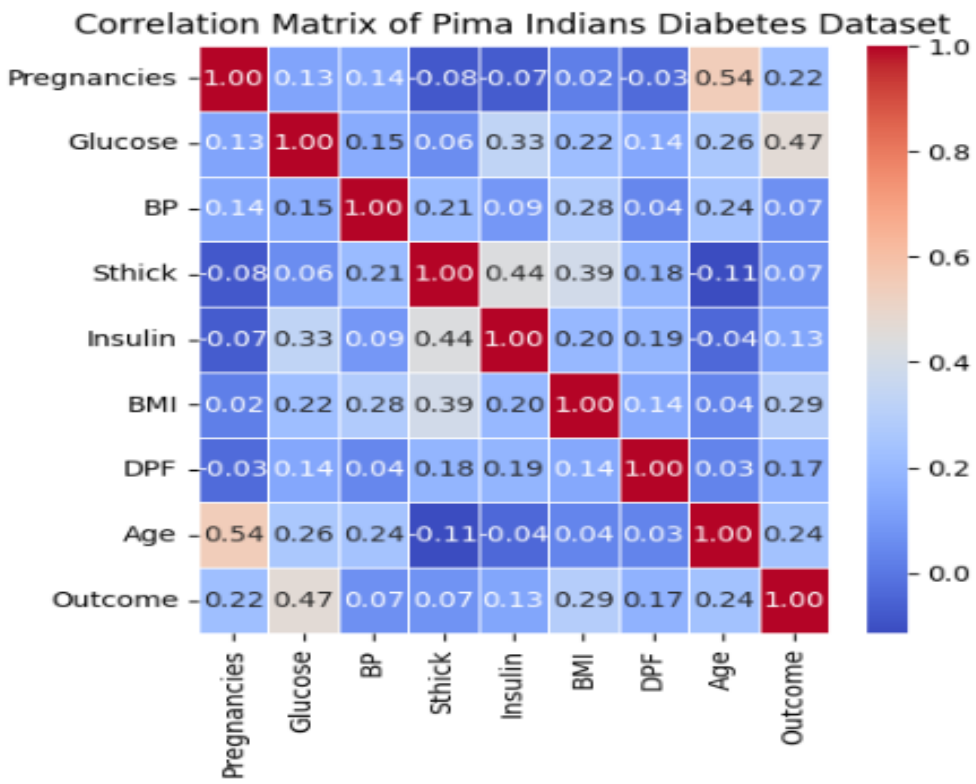


Fig. 3 Correlation matrix.

The variable pairs having the highest correlation are displayed in the correlation matrix. There is a fairly strong association between the target variables and age, pregnancy, BMI, insulin, and glucose

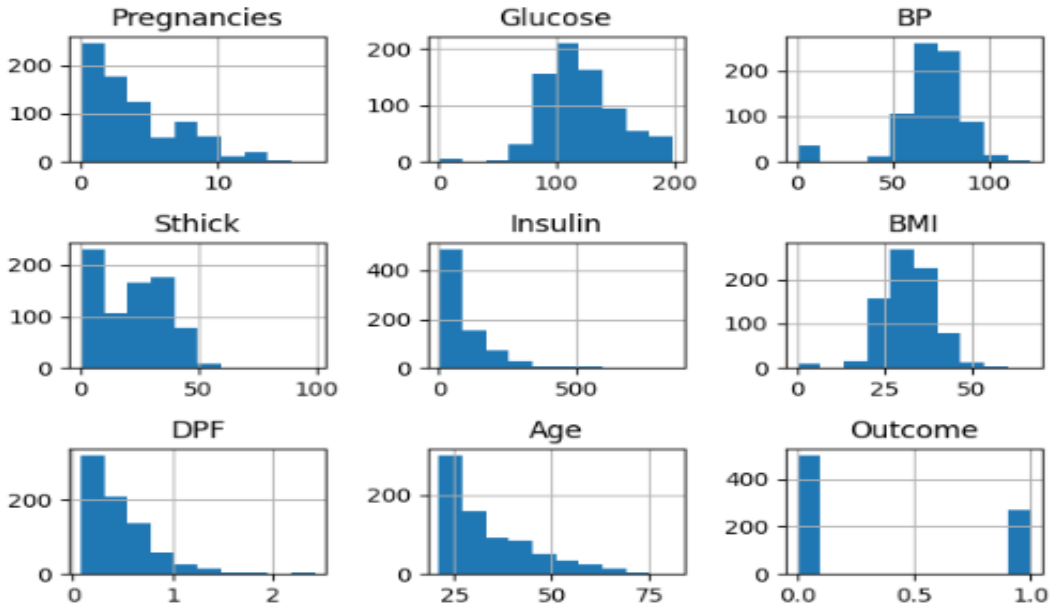


Fig. 4 Plot distribution for each attribute

The columns for Age and Insulin are clearly slanted to the right. Because of this, a normalizing procedure was required before to the modeling phase. Many individuals in the dataset were in the 20–40 age range; the majority had blood pressure between 50 and 100 mmHg and insulin levels of 0. The majority of individuals were classified as prediabetes patients and had blood glucose levels between 140 and 199 mg/dL. While healthy persons should have a BMI between 18.5-24.9, the readings ranged from 20 to 50. According to this dataset, a large number of persons were overweight.

Let us perform preprocessing

Handling Missing Value:As seen in Fig. 5(a), The variables with the highest percentage of missing data are insulin (374 data, or 48.7%), skin thickness (227 data, or 29.56%), blood pressure (35 data, or 4.56%), BMI (11 data, or 1.43%), and glucose (5 data, or 0.65%).

All missing data values were substituted with the median value of each variable to eliminate any remaining empty variable values. After removing missing values the result was shown as below.

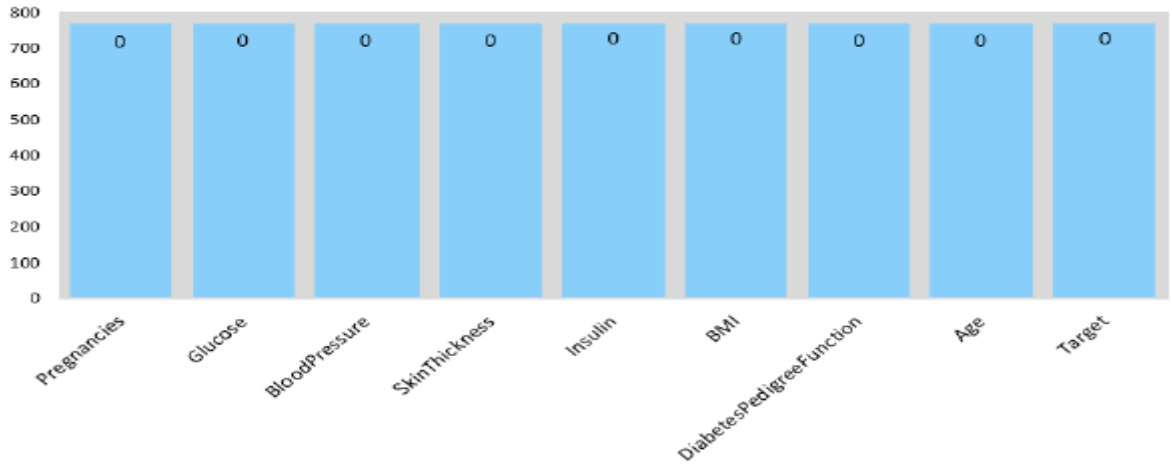


Fig: 5 After removing missing values

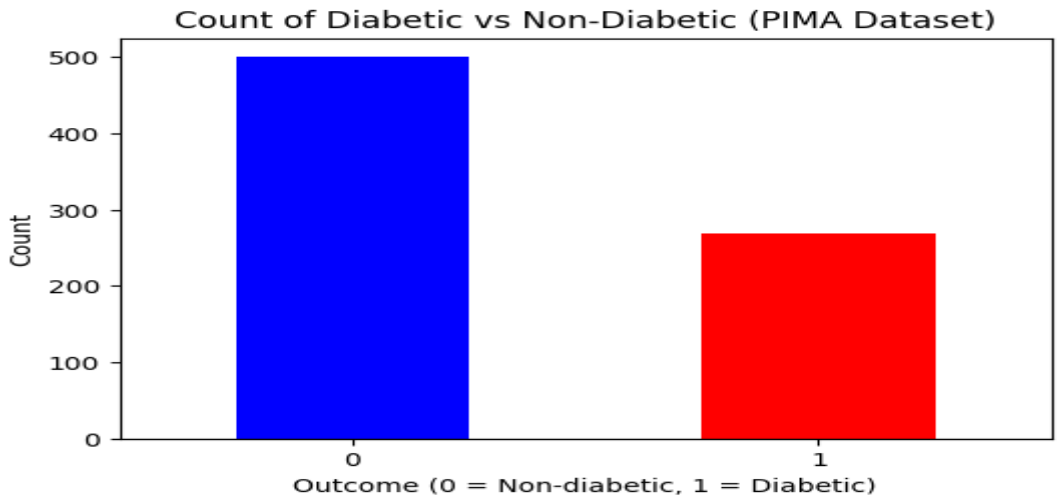


Fig. 6 Imbalanced dataset

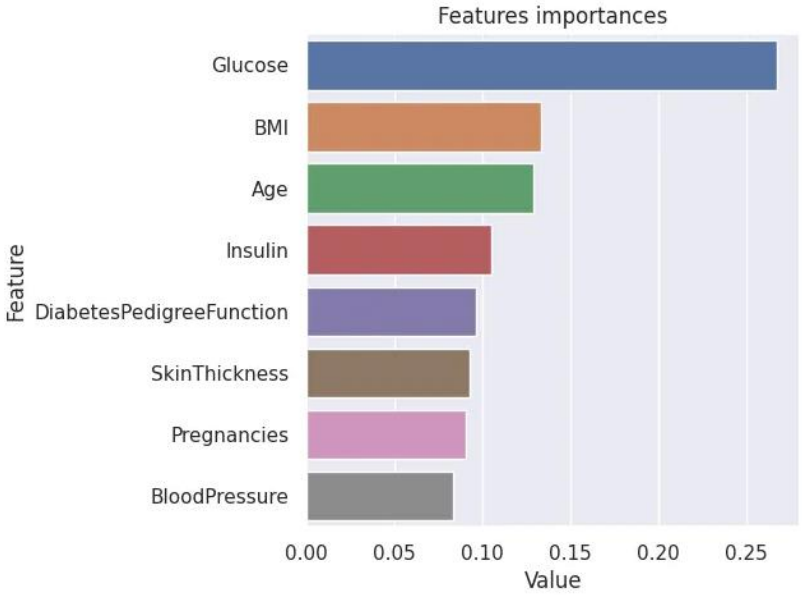


Fig:7 Features importance

Imbalanced Dataset: Fig. 6 showed unbalanced dataset and There are 268 patients (34.9%) with diabetes and 500 patients (65.1%) without the disease. To counteract the unbalanced data, the SMOTE approach was employed. After the preprocessing stage, the next step was modeling using logistic regression. The data was first divided into training and test sets. The test data to training data ratio was 70:30.

Model evaluation was carried out after the model was formed. Table III shows the evaluation results of the model in confusion matrix format.

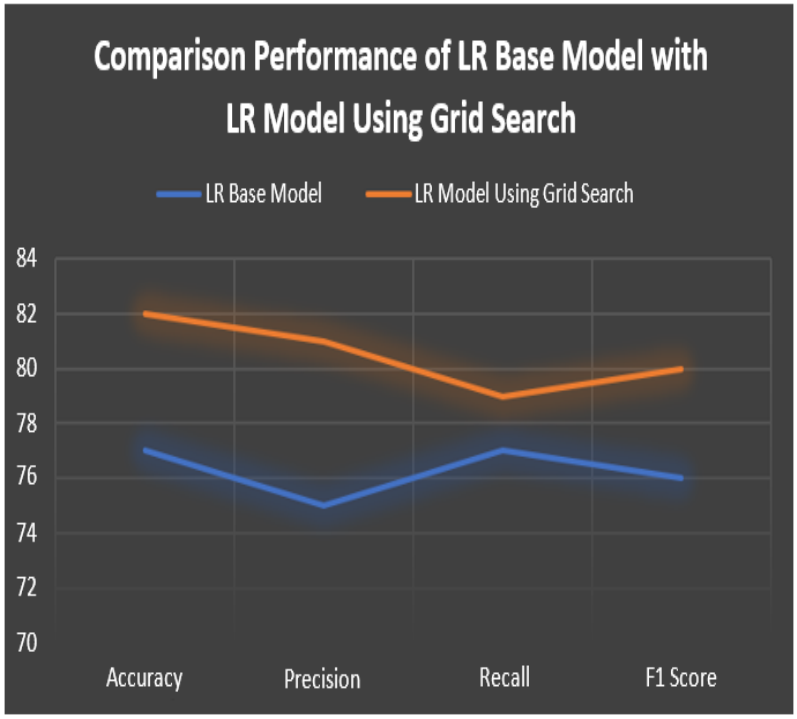
Table 2 Confusion matric for Logistic Regression

	Predicted Positive	Predicted Negative
Actual Positive(Non Diabetic)	116	35
Actual Negative(Diabetic)	18	62

172 data were in the proper classification (true positive and true negative), consisting of 116 data predicted to be diabetic and, in fact, had diabetes and 62 people were expected

to be non-diabetes and, in fact, were non-diabetes. A total of 53 other data were false positive and false negative, namely, 35 non-diabetes people were predicted to have diabetes, and eighteen people with diabetes were expected to be non-diabetes. Table 2 shows model performance measurement based on accuracy, precision, recall, and F1-score values. The precision value obtained for the non-diabetes class was 87%, and for diabetes was 64%, with an average precision value of 75%.

The average recall value was 77% for those without diabetes and 78% for those with the disease. Precision and recall were not significantly different from the F1-score. With an average of 76%, the F1-score for diabetes was 70% and for non-diabetes it was 81%. 77% was the accuracy value Grid Search Technique for Hyperparameter Tuning was one of the parameters utilized to enhance the model's performance. A sample of a Python script that uses a grid search method to choose the optimal parameters is shown in Fig. 8. The best penalty regulation was determined to be L2, with a C value of 0.6158 being the most ideal.



The impact of employing hyperparameter tuning was measured by comparing the model's performance before and after the grid search technique was applied. The average accuracy, precision, recall, and F1-score for the basic model were 77%, 75%, and 77%, respectively. In contrast, the improved model's accuracy value rose to 82%, precision to 81%, recall to 79%, and F1-score to 80%.

5. Conclusion

This study successfully implemented the logistic regression algorithm in predicting diabetes with a good accuracy. Understanding the data was done through data exploration and research to analyze pairs of variables that had a reasonably strong correlation to the determination of the target value through visualization techniques in the form of distributions and scatter plots. The grid search technique was used to enhance the performance of the logistic regression algorithm's fundamental model. After applying hyperparameter adjustment, the model's performance improved, according to an evaluation of the model using the confusion matrix. Thus, the study's experimental

findings demonstrate that one of the most effective methods for creating prediction models is the logistic regression algorithm combined with the grid search methodology. By merging logistic regression algorithms with additional classification algorithms like random forests, support vector machines, and k-nearest neighbor with ensemble techniques, future research can employ deep learning algorithms on larger datasets.

Reference

1. P. Chandra Sen, M. Hajra, and M. Ghosh "Supervised Classification Algorithms in Machine Learning: A Survey and Review" Advances in Intelligent Systems and Computing, vol. 937. 2020.
2. Ibrahim and Abdulazeez Adnan "The role of machine learning algorithms for diagnosing diseases" Journal of Applied Science and Technology Trends, vol. 2, no. 1, pp. 10–19, 2021.
3. H. Johns, J. Bernhardt, and L. Churilov "Distance-based Classification and Regression Trees for the analysis of complex predictors in health and medical research" Statistical Methods in Medical Research, vol. 30, no. 9, 2021, doi:10.1177/09622802211032712.
4. A. Lestari "Increasing Accuracy of C4 . 5 Algorithm Using Information Gain Ratio and Adaboost for Classification of Chronic Kidney Disease" Journal of Soft Computing Exploration, 2020.157
5. K. Chowdhury, D. Chaudhuri, and A. K. Pal "An entropy-based initialization method of K-means clustering on the optimal number of clusters" Neural Computing and Applications, vol. 33, no. 12, 2021, doi: 10.1007/s00521-020-05471-9.
6. H. Puri, J. Chaudhary, K. R. Raghavendra, R. Mantri, and K. Bingi "Prediction of Heart Stroke Using Support Vector Machine Algorithm" 2021. doi:10.1109/ICSCC51209.2021.9528241.158
7. A. F. Marquand and S. M. Kia, "Linear methods for classification" Machine Learning. Academic Press, 2020. 83-100.
8. Diabetes Federation International and IDF, IDF Diabetes Atlas 2019, 9th Editio. 2019
9. H. Puri, J. Chaudhary, K. R. Raghavendra, R. Mantri, and K. Bingi "Prediction of Heart Stroke Using Support Vector Machine Algorithm" 2021. doi:10.1109/ICSCC51209.2021.9528241.
10. Evwiekpaefe, A. E., Abdulkadir, N., Nigerian Defence Academy, & Nigerian Defence Academy. (2023). A predictive model for diabetes mellitus using machine learning techniques (A study in Nigeria). The African Journal of Information Systems, 1–1. https://digitalcommons.kennesaw.edu/ajis/vol15/iss1/1
11. Okikiola, F. M., Adewale, O. S., & Obe, O. O. (2023). A DIABETES PREDICTION CLASSIFIER MODEL USING NAIVE BAYES ALGORITHM. FUDMA Journal of Sciences, 7(1), 253–260. https://doi.org/10.33003/fjs-2023-0701-1301
12. Hasan, M. K., Alam, M. A., Das, D., Hossain, E., & Hasan, M. (2020). Diabetes prediction using ensembling of different machine learning classifiers. IEEE Access, 8, 76516–76531. https://doi.org/10.1109/access.2020.2989857
13. Lai, H., Huang, H., Keshavjee, K., Guergachi, A., & Gao, X. (2019). Predictive models for diabetes mellitus using machine learning techniques. BMC Endocrine Disorders, 19(1). https://doi.org/10.1186/s12902-019-0436-6

References

1. Global Burden of Disease Collaborative Network, Global Burden of Disease Study 2021. Results, Institute for Health Metrics and Evaluation, 2024. https://vizhub.healthdata.org/gbd-results/
2. T. Panch, P. Szolovits, R. Atun, Artificial intelligence, machine learning, and health systems, Journal of Global Health, 8 (2018) 020303. https://doi.org/10.7189/jogh.08.020303
3. S. Uddin, A. Khan, M.E. Hossain, M. Rahman, et al., Comparing different supervised machine learning algorithms for disease prediction, BMC Medical Informatics and Decision Making, 19 (2019) 1–16.https://doi.org/10.1186/s12911-019-1004-8
4. A. Panesar, Machine Learning and AI for Healthcare, Springer, 2019. https://doi.org/10.1007/978-1-4842-3799-1
5. E. Erlin, Y.N. Marlim, Junadhi, L. Suryati, N. Agustina, Early Detection of Diabetes Using Machine Learning with Logistic Regression Algorithm, Jurnal Nasional Teknik Elektro dan Teknologi Informasi, 11 (2022) 88–96. https://doi.org/10.22146/jnteti.v11i2.3586
6. Smith J. Pima Indians Diabetes Database. Kaggle; 2016. Available from: https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database [Accessed 8 Sep 2025].

***

ABOUT EMBAR PUBLISHERS

Embar Publishers is an open-access, international research based publishing house committed to providing a 'peer reviewed' platform to outstanding researchers and scientists to exhibit their findings for the furtherance of society to provoke debate and provide an educational forum. We are committed about working with the global researcher community to promote open scholarly research to the world. With the help of our academic Editors, based in institutions around the globe, we are able to focus on serving our authors while preserving robust publishing standards and editorial integrity. We are committed to continual innovation to better support the needs of our communities, ensuring the integrity of the research we publish, and championing the benefits of open research.

Our Journals

1. Research Journal of Education , linguistic and Islamic Culture - 2945-4174
2. Research Journal of Education and Advanced Literature – 2945-395X
3. Research Journal of Humanities and Cultural Studies - 2945-4077
4. Research Journal of Arts and Sports Education - 2945-4042
5. Research Journal of Multidisciplinary Engineering Technologies - 2945-4158
6. Research Journal of Economics and Business Management - 2945-3941
7. Research Journal of Multidisciplinary Engineering Technologies - 2945-4166
8. Research Journal of Health, Food and Life Sciences - 2945-414X
9. Research Journal of Agriculture and Veterinary Sciences - 2945-4336
10. Research Journal of Applied Medical Sciences - 2945-4131
11. Research Journal of Surgery - 2945-4328
12. Research Journal of Medicine and Pharmacy - 2945-431X
13. Research Journal of Physics, Mathematics and Statistics - 2945-4360