International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 03 | March - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Diabetes Prediction Using Machine Learning Algorithm

Mrs. Sindhu Jain A M, Abdul Faraz, Abhishek K, Jeevitha, Keerthana D Y
Information Science and Engineering, Visvesvaraya Technological University
Sindhujainam96@gmail.com
Abdul360faraz@gmail.com
Abhishekkumarkr02@gmail.com
jeevithachaluvaraj@gmail.com
keerthanady014@gmail.com

Abstract- Diabetes is a current and serious health condition affecting millions worldwide. It can affect from multiple factors, including inheritable predilection, life habits, poor diet, and high blood pressure.However, diabetes significantly increases the liability of complications similar as cardiovascular conditions, order failure, If left unmanaged. The ideal of this design is to develop a machine literacy- grounded vaticination model that evaluates colourful health pointers similar as blood glucose situations, BMI, age, and blood pressure to assess an existent's threat of developing diabetes. By using machine literacy ways, the system enables early discovery and timely intervention, making it accessible through web- grounded or mobile platforms to enhance usability and scalability.

I. INTRODUCTION

Diabetes is a habitual condition characterized by elevated blood glucose situations, posing a growing health challenge worldwide. The complaint is associated with severe complications, including cardiovascular conditions, strokes, order diseases, and vision impairment. Beforehand discovery and visionary operation are pivotal to reducing adverse health issues.

Recent advancements in machine literacy have revolutionized prophetic analytics in healthcare, offering important tools for assaying large datasets. These algorithms descry patterns that traditional statistical styles may overlook. This study aims to apply machine literacy ways to prognosticate diabetes threat grounded on case-specific attributes similar as BMI, age, and blood pressure. The developed model will help healthcare professionals in relating high-threat individualities and enforcing timely preventative measures.

II. LITERATURE REVIEW

Several experimenters have developed machine literacy- grounded vaticination models for diabetes discovery using colourful ways. For case, studies have employed decision trees, Naïve Bayes classifiers, and support vector machines to dissect diabetes datasets. Experimenters similar as Kumar et al. 2015) enforced a system integrating Hadoop and MapReduce, which successfully distributed diabetes types and associated threat factors. also, Iyer (2015) explored hidden patterns within diabetes datasets using bracket algorithms, demonstrating the effectiveness of Naïve Bayes and Decision Trees.

Rajesh and Sangeetha (2012) employed the C4.5 decision tree algorithm for bracket, while Kahramanli and Allahverdi (2008) combined artificial neural networks with fuzzy sense to enhance vaticination delicacy. Also, Patil et al. (2010) introduced a mongrel approach using K- means clustering followed by bracket algorithms to ameliorate vaticination trustability. These studies emphasize the significance of using machine literacy in diabetes vaticination, inspiring the proposed model in this exploration.

III. PROBLEM STATEMENT

The rising prevalence of diabetes necessitates the development of effective discovery systems able of relating at-threat individualities. Numerous being styles calculate on clinical data, which may not always be available for large populations. Likewise, conventional individual ways warrant the availability and delicacy needed for wide perpetration. This study aims to bridge this gap by creating an accurate and accessible machine literacy- grounded system for diabetes vaticination.

IV. PROPOSED SYSTEM

768 patient records with eight essential characteristics make up the dataset used.

Sl.No	Attributes
1.	Pregnancies
2.	Glucose
3.	Blood Pressure
4.	Skin Thickness
5.	Insulin
6.	BMI(Body Mass Index
7.	DiabetesPedegreeFunction
8.	Age

Certain aspects of the dataset, such BMI and glucose, have zero values, which means that there are missing data points that need to be fixed during pre-processing. With 268 cases of diabetes and 500 non-diabetic patients, the dataset is somewhat unbalanced. To increase prediction accuracy, the model will employ classification and ensemble approaches.

V. METHODOLOGY

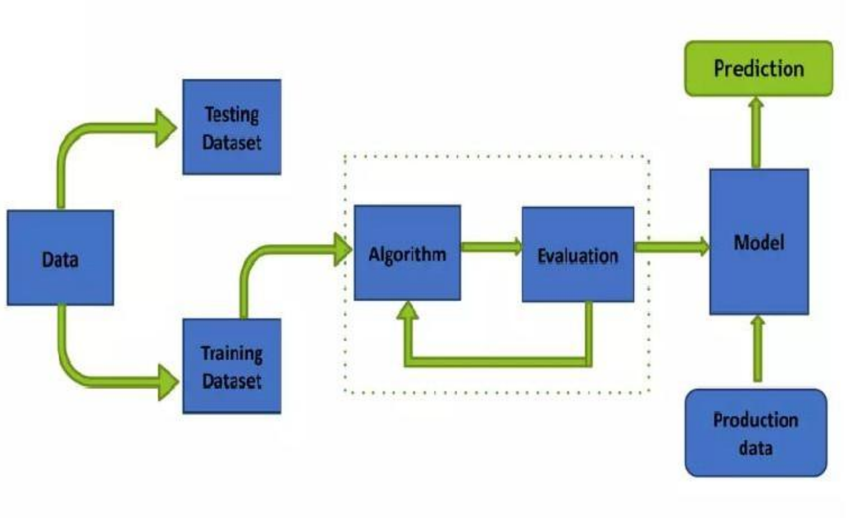

Fig 1: Methodology of Diabetes Prediction

The methodology consists of the following way

1. Data Collection Dataset: Use the PIMA Indian Diabetes Dataset or analogous.
   Features include glucose position, BMI, age, blood pressure, Insulin.
2. Data Preprocessing: Is critical for icing the quality of the input which directly affects the performance of the machine literacy model.
3. Model Selection: Estimate the following machine literacy models.
   Logistic Retrogression, Decision Trees Random Forest, Support Vector Machines
4. Model Training and Testing: Split data into training and testing sets (e.g., 80- 20). Estimate models using criteria like delicacy, perfection, recall, and F1- score.

Volume: 09 Issue: 03 | March - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

5. Model Development: Once the best- performing model is named. It's stationed for real- world operation.

VI. EXPERIMENTAL RESULT

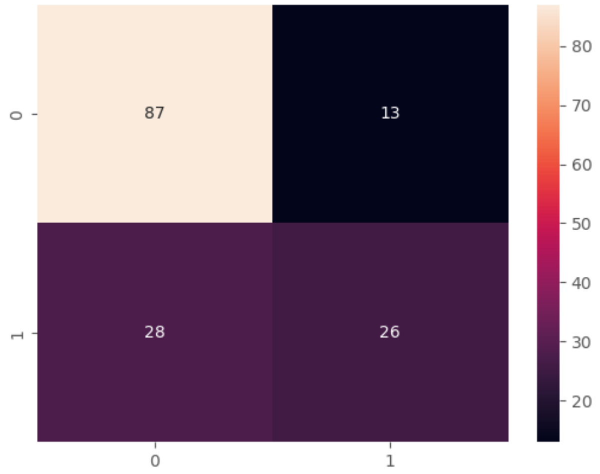

Fig 2: Confusion Matrix Visualization

This representation of a confusion matrix is frequently used to assess how well a classification model performs. This is what it stands for.

A. Structure

1. The row represents the actual class labels.
2. The column represents the predicted class labels.
3. Each cell contains the count of instance falling into the category.

B. Interpreting the values

The numbers in the matrix are as follows:

1. 87 (Top-left, True Negative, or TN): The model correctly predicted Class 0 (negative class).
2. 13. The model predicted class 1 when it was actually class 0 (False Positive, FP, top-right).
3. 28 (False Negative, FN, bottom-left): Class 1 was predicted by the model, while class 0 was actually predicted.
4. 26 (bottom-right, True Positive, or TP): The model correctly predicted Class 1 (positive class).

C. Performance Analysis

1. The model correctly predicted 87 + 26 = 113 occurrences.
2. The model classified things incorrectly in 13 + 28 = 41 cases.
3. Since the model was unable to identify real positive cases, the false negatives (28) might pose a serious issue if class 1 indicates a severe situation (disease diagnosis, for instance).

1. Accuracy

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Substituting the values:

$$\text{Accuracy} = \frac{87 + 26}{87 + 26 + 13 + 28} = \frac{113}{154} \approx 0.7338\ (73.38\%)$$

---

2. Precision (for class 1, positive class)

$$\text{Precision} = \frac{TP}{TP + FP}$$

Substituting the values:

$$\text{Precision} = \frac{26}{26 + 13} = \frac{26}{39} \approx 0.6667\ (66.67\%)$$

D. Summary of Metrics

1. Accuracy: 73.38%
2. Precision(for positive class): 66.67%

VII. CONCLUSION

The design successfully developed a machine literacy- grounded diabetes vaticination system that provides high delicacy in relating individualities at threat. By using bracket algorithms, point selection ways, and rigorous model evaluation, the system offers a dependable and scalable result for diabetes threat assessment. Unborn advancements may include integrating real- time case data from electronic health records and wearable bias to ameliorate model robustness.

VIII. REFERENCES

[1] Gauri D. Kalyankar, Shivananda R. Poojara, Nagaraj V. Dharwadkar.' Prophetic Analysis of Diabetic Patient Data Using Machine literacy and Hadoop.' International Conference on I- SMAC, 2017.

[2] Ayush Anand, Divya Shakti.' Diabetes vaticination Using Individual life pointers.' 1st International Conference on Next Generation Computing Technologies, 2015.

[3] B. Nithya, Dr. V. Ilango.' Prophetic Analytics in Health Care Using Machine Learning Tools and ways.' International Conference on Intelligent Computing and Control Systems, 2017.

[4] Dr. Saravana Kumar N M, Eswari T, Sampath P, Lavanya S.' Prophetic Methodology for Diabetic Data Analysis in Big Data.' 2nd International Symposium on Big Data and Cloud Computing, 2015.

[5] Aiswarya Iyer, S. Jeyalatha, Ronak Sumbaly.' opinion of Diabetes Using Bracket Mining ways.' International Journal of Data Mining & Knowledge Management Process, 2015.

[6] Mani Butwall, Shraddha Kumar.' A Data Mining Approach for the opinion of Diabetes Mellitus using Random Forest Classifier.' International Journal of Computer Applications, 2015.