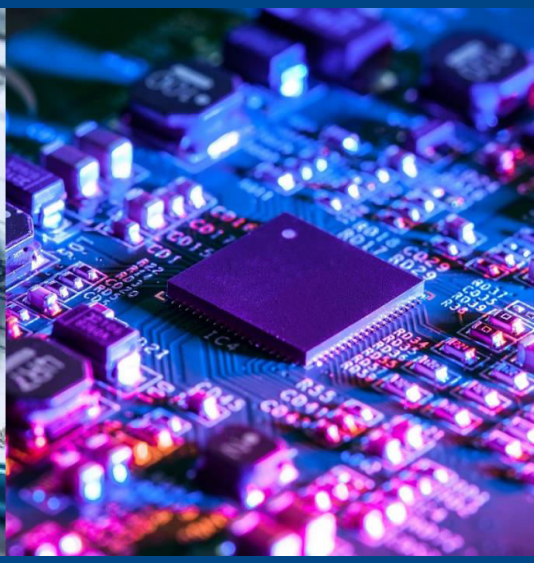
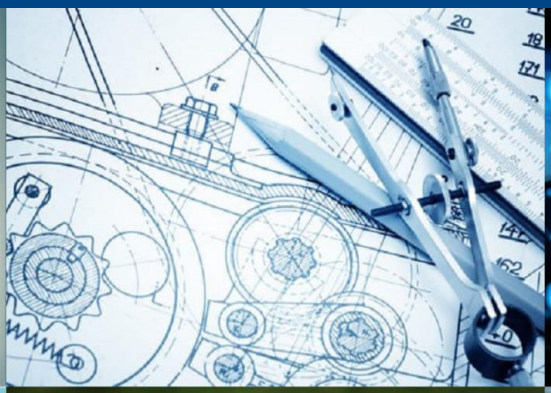
ISSN: 2582-7219

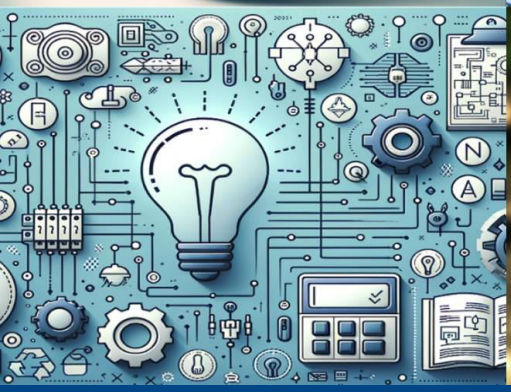




International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.206 **Volume 8, Issue 4, April 2025**

© 2025 IJMRSET | Volume 8, Issue 4, April 2025| DOI:10.15680/IJMRSET.2025.0804235

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Diabetes Predicting System Using Machine Learning

Bagwan A.A[1], Makandar S.M[2], Hulle Y.B[3], Talikoti V.V[4], Jadhav A.R[5], Prof. Pawar V.D[6], Prof. Chougule V.V[7]

Department of Computer Engineering, SVSMD's KKI Polytechnic College, Akkalkot, India[1-7]

ABSTRACT: Diabetes mellitus, characterized by chronic hyperglycemia, poses a significant global health challenge. Early detection and prediction are crucial for effective management and mitigation of long-term complications. This study investigates the application of machine learning (ML) algorithms for diabetes prediction, aiming to enhance the accuracy and efficiency of diagnostic processes. We analyze the performance of Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Logistic Regression, and Random Forest classifiers using a well-known diabetes dataset. Performance metrics, including Recall, F1-Measure, Precision, and Accuracy, are employed to evaluate the predictive capabilities of each algorithm. The results contribute to the growing body of research on ML-driven diabetes prediction, offering insights into the strengths and limitations of different classification techniques.

KEYWORDS: Diabetes Mellitus, Machine Learning, Predictive Analytics, Classification Algorithms, Healthcare, Data Mining

I. INTRODUCTION

In this day and age, one of the most notorious diseases to have taken the world by storm is Diabetes, which is a disease which causes an increase in blood glucose levels as a result of the absence or low levels of insulin. Due to the many criteria to be taken into consideration for an individual to harbor this disease, it's detection and prediction might be tedious or sometimes inconclusive. Nevertheless, it isn't impossible to detect it, even at an early stage .

Diabetes is increasing day by day in the world because of environmental, genetic factors. The numbers are rising rapidly due to several factors which includes unhealthy foods, physical inactivity and many more. Diabetes is a hormonal disorder in which the inability of the body to produce insulin causes the metabolism of sugar in the body to be abnormal, thereby, raising the blood glucose levels in the body of a particular individual. Intense hunger, thirst and frequent urination are some of the observable characteristics. Certain risk factors such as age, BMI, Glucose Levels, Blood Pressure, etc., play an important role to the contribution of the disease.

Diabetes is a very familiar word in the present world and crucial challenges in both developed and developing countries . The insulin hormone in the body produced by the pancreas allows glucose to pass from the food into the bloodstream. The lack of that hormone due to malfunctioning of the pancreas forms diabetes which can result in coma, renal and retinal failure, pathological destruction of pancreatic beta cells, cardiovascular dysfunction, cerebral vascular dysfunction, peripheral vascular diseases, sexual dysfunction, joint failure, weight loss, ulcer, and pathogenic effects on immunity.

Diabetes is the third leading cause of death following diseases of heart and cancer. But with the rise of Machine Learning approaches, we have the ability to find a solution to this issue. The aim of Machine Learning and Data Mining is to extract knowledge from information stored in dataset and generate clear and understandable description of patterns. We are going to develop a Diabetes Diagnosis system using Machine Learning which has the ability to predict whether the patient has diabetes or not. Furthermore, predicting the disease early leads to treating the patients before it becomes critical. Machine Learning and Data mining has the ability to extract hiddenknowledge from a huge amount of diabetes-related data. This paper reviewed and analyzed the current studies on classification of Diabetes. Furthermore, the study has developed a classification model for diabetes using decision tree, Naïve Bayes, Support vector machine and k nearest neighbor Algorithm. The classification model is based on a dataset of 15000 cases collected from different National Institute of Diabetes and Digestive and Kidney Diseases. The results of the Naïve Bayes can be used by medical specialist to classify and diagnose diabetic patients. These results help the medical doctors in the classification process of diabetes. This study follows different machine learning algorithms to predict diabetes disease

© 2025 IJMRSET | Volume 8, Issue 4, April 2025| DOI:10.15680/IJMRSET.2025.0804235

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

at an early stage. Such as, KNN, Naïve Bayes, Decision Tree, and Support Vector machine to predict this chronic disease at an early stage for safe human life.

II. EXISTING SYSTEM

Diabetes Prediction System by PYRESEARCH:
This system uses machine learning algorithms to predict diabetes based on key health indicators such as glucose levels, blood pressure, and BMI. It features a user-friendly interface and provides real-time processing and detailed reporting

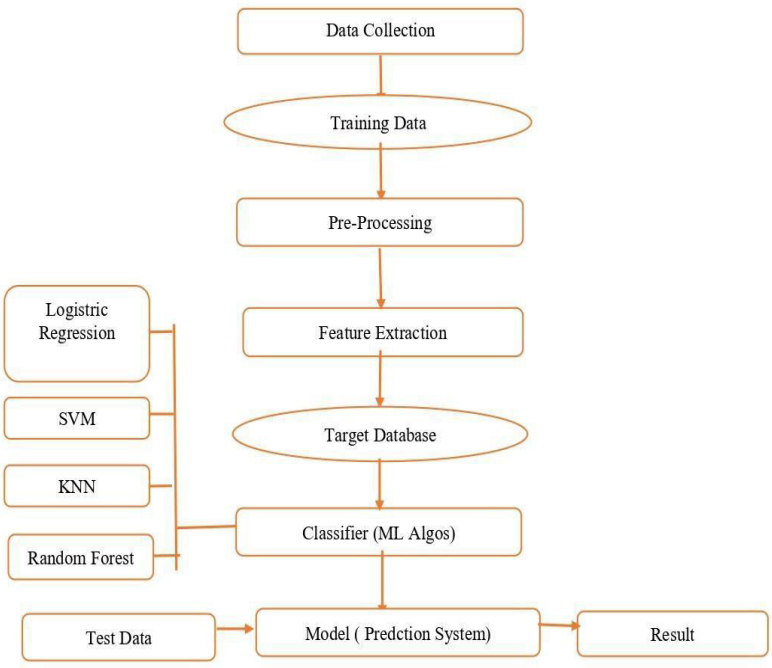
III. NEED OF THIS PROJECT

This research work aims to analyze the Diabetes dataset, design, and implement a Diabetes prediction and recommendation system utilizing machine learning classification techniques. The specific objectives of this project work are:

1. To review existing literature along the area of diabetes diagnosis and prediction.
2. Design and develop a model using machine learning techniques.
3. To analyze the Diabetes dataset and use Support Vector Machine and Random forest algorithms to develop a prediction engine.
4. To identify and discuss the benefits of the designed system along with effective applications.

IV. PROPOSED SYSTEM

The proposed model is introduced to overcome all the disadvantages that arises in the existing system.
This system will increase the accuracy of the Supervised classification results by classifying the data based on the social network mental disorders and others using Decision tree classification algorithm.
It enhances the performance of the overall classification results.
Apply data mining techniques to the dataset to investigate if ML & DL techniques can achieve equivalent (or better) results in identifying suitable treatments as that achieved in the diagnosis.



© 2025 IJMRSET | Volume 8, Issue 4, April 2025| DOI:10.15680/IJMRSET.2025.0804235

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

V. METHODOLOGY

- Research Design:This study employs a quantitative research design to investigate the effectiveness of machine learning algorithms for diabetes prediction. The methodology involves four main phases:
- Dataset:The dataset used in this study was obtained from the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK). The dataset contains 9 attributes, including:
  * Pregnancies
  * Glucose
  * Blood Pressure
  * Skin Thickness
  * Insulin
  * BMI
  * Diabetes Pedigree Function
  * Age
  * Outcome (diabetes positive or negative)

  The dataset comprises 768 records of female patients.

- Data Pre-Processing:Data preprocessing is a crucial step in machine learning to ensure the quality and accuracy of the models. The following preprocessing steps were applied:
- Handling Missing Values: Missing values, if any, were addressed using appropriate imputation techniques (e.g., mean imputation, median imputation).
- Data Normalization: The data was normalized to a specific range (e.g., 0 to 1) to prevent features with larger scales from dominating the learning process.
- Data Splitting: The preprocessed data was split into training and testing sets. The training set (e.g., 80% of the data) was used to train the machine learning models, and the testing set (e.g., 20% of the data) was used to evaluate their performance.
- Machine Learning Algorithms:The following machine learning algorithms were implemented for diabetes prediction:
  K-Nearest Neighbors (KNN):
  Logistic Regression:
  Random Forest:
  Support Vector Machine (SVM):
- Performance Evaluation Metrics:The performance of the machine learning algorithms was evaluated using the following metrics:
  Accuracy: The proportion of correctly classified instances.
  Precision: The proportion of correctly predicted positive cases out of all cases predicted as positive.
  Recall (Sensitivity): The proportion of correctly predicted positive cases out of all actual positive cases.
  F1-Measure: The harmonic mean of precision and recall.
  Confusion Matrix: A table that summarizes the performance of a classification model.

VI. CONCLUSION

This project compared the effectiveness of Logistic Regression with other linear classifiers (SVM, KNN, Random Forest) for diabetes prediction. Logistic Regression outperformed others with an accuracy of 83%. The proposed approach utilized ensemble learning and classification methods, enabling early predictions and informed decisions for healthcare professionals.The project's goal was to design and implement diabetes prediction using machine learning, which was achieved successfully. The experimental results can assist healthcare professionals in taking early predictions and making informed decisions to cure diabetes and save human lives.The proposed approach demonstrated the potential of machine learning in healthcare, particularly in disease diagnosis and treatment. By leveraging these methods, healthcare professionals can improve patient outcomes and reduce the risk of complications.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

VII. FUTURE WORK

In the future, we plan to collaborate with hospitals or medical institutes to create a comprehensive diabetes dataset, aiming to improve our model's performance. We'll explore additional Machine Learning and Deep Learning models to achieve better results. To further enhance our model, it's crucial to gather and incorporate real-time data from hospitals, allowing for continuous training and development.With a larger dataset, we'll be able to apply more advanced data mining techniques, expanding our predictive capabilities. By developing a robust framework and quality approach, we can improve the accuracy of our predictions and reduce the risk of complications associated with diabetes. This will ultimately help minimize the growing burden of diabetes and improve patient outcomes.

REFERENCES

[1] Arwatki Chen Lyngdoh, Nurul Amin Choudhury, Soumen Moulik, “Diabetes Disease Prediction Using Machine Learning Algorithms”, IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES) 2020, DOI: 10.1109/IECBES48179.2021.9398759 .

[2] Mitushi Soni, Dr. Sunita Varma " Diabetes Prediction using Machine Learning Techniques" , Journal of Engineering Research & Technology (IJERT) 2020,

[3] Sivaranjani S, Ananya S, Aravinth J, Karthika R, " Diabetes Prediction using Machine Learning Algorithms with Feature Selection and Dimensionality Reduction”, 7th International Conference on Advanced Computing and Communication Systems (ICACCS), 2021 DOI: 10.1109/ICACCS51430.2021.9441935 .

[4] Shejal Kale, Priti Rahane, Mansi Ghumare, Snehal PatilB “Diabetes Prediction Using Different Machine Learning Approaches” IJSDR | Volume 7 Issue 5 , 2022.

[5] Ashwini r, s m aiesha afshin, kavya v, deepthi raj “diabetes prediction using machine learning” ijrti | volume 7, issue 7 | issn: 2456-3315, 2022 .

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com