International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 05 | May – 2025 DOI: 10.55041/ISJEM03457
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

DIABETES PREDICTION USING MACHINE LEARNING

Mr.A.MOHAMED ARAFATH, Ms.S.MONIKA, Mr.RITHUL JANARDHAN, Ms.R.SANDHYA

Mrs.M.MOHANAPRIYA (MENTOR)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

SRI SHAKTHI INSTITUTE OF ENGINEERING AND TECHNOLOGY

COIMBATORE

Abstract - In an era where digital healthcare solutions are becoming increasingly essential, the "Diabetes Prediction Website" is designed to revolutionize the way individuals assess their risk of diabetes through intelligent, user-friendly online tools. This project focuses on building a responsive and accurate web-based system that empowers users to predict their likelihood of developing diabetes based on medical and lifestyle inputs. Leveraging the power of machine learning, data science, and modern web technologies, the system ensures high accuracy and personalized results in real-time.

The Diabetes Prediction Website enables users to input health metrics such as age, BMI, blood pressure, glucose level, insulin levels, and more, to receive immediate risk assessment outcomes. Key components of the system include data preprocessing, model training using algorithms like Logistic Regression and Random Forest, and an interactive frontend interface. The platform is backed by a secure database and scalable backend infrastructure, ensuring smooth operation and data privacy. Highly adaptable, this system can be integrated into telehealth platforms, fitness applications, or digital clinics. By simplifying risk assessment, promoting awareness, and enabling early intervention, the Diabetes Prediction Website reflects a modern, data-driven approach to preventive healthcare. Ultimately, this solution aims to foster proactive health management and improve quality of life through smart diagnostics.

***Key Words*:** risk assessment, Real-time prediction, Data preprocessing, Smart diagnostics, Machine learning, Accuracy, Fitness applications

1.INTRODUCTION

Diabetes Prediction Website is a modern, data-driven tool designed to help users assess their risk of diabetes quickly and accurately. Unlike traditional diagnostic methods that can be time-consuming and inaccessible, this platform uses machine learning and real-world health data to deliver instant risk predictions based on inputs like age, BMI, glucose, blood pressure, and insulin levels.

The system is built for scalability, featuring a user-friendly interface, secure data handling, and integration capabilities with telehealth platforms and wellness apps. It empowers individuals to monitor their health proactively while reducing pressure on healthcare systems. This project promotes early intervention, informed decision-making, and greater accessibility in diabetes management.

2. Body of Paper

2.1 Machine Learning Algorithms

Logistic Regression: A statistical method used for binary classification problems, such as predicting whether a patient is diabetic or not based on input features.

Decision Trees: A model that uses a tree-like graph of decisions and their possible consequences, useful for classification and regression tasks.

Random Forest: An ensemble learning method that constructs multiple decision trees during training and outputs the mode of their predictions, improving accuracy.

Gradient Boosting: An ensemble technique that builds models sequentially, where each new model corrects errors made by the previous ones, enhancing prediction performance.

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 05 | May – 2025 DOI: 10.55041/ISJEM03457
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

2.2 Real-Time Data Processing Algorithms

Kalman Filters: Used for estimating the state of a dynamic system from a series of incomplete and noisy measurements, useful for real-time health monitoring.
Time Series Analysis: Techniques for analyzing time-ordered data points, which can be applied to monitor health metrics over time.

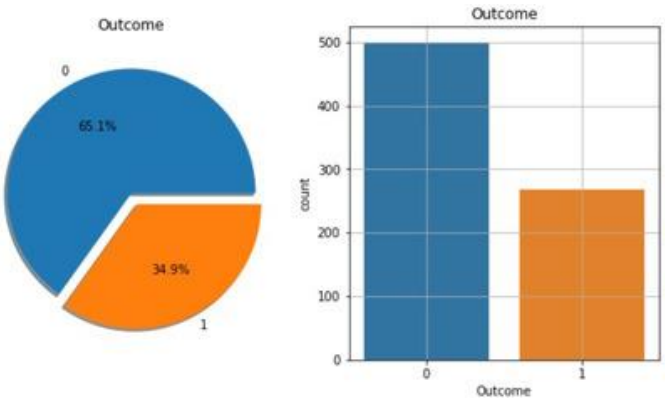

Fig -1: Chart

3. METHODOLOGY

Traditionally, diabetes diagnosis has relied on manual methods conducted by healthcare professionals, involving clinical tests, physical examinations, and a review of the patient's medical history. This process typically requires in-person visits to healthcare facilities, blood tests, and medical consultations, making it both time-consuming and resource-intensive. For individuals living in remote or underserved areas, access to these services can be limited, leading to delays in diagnosis and missed opportunities for early intervention—especially in cases where patients are asymptomatic of Risk.

In an effort to provide basic risk assessment outside clinical settings, some online tools have adopted rule-based calculators. These systems use predefined scoring methods based on factors like age, weight, and family history. However, such tools are static in nature and lack the capability to adapt to the unique variations in individual health profiles. They fail to capture the complex, non-linear relationships between health parameters and often yield low predictive accuracy, resulting in a poor substitute for intelligent diagnostics.

Another major limitation of conventional approaches is the absence of real-time prediction and automation. In most cases, users cannot receive immediate feedback based on self-reported data, which delays preventive measures and reduces overall engagement with personal health management. Without instant insights, users are less likely to take timely action, undermining the goal of proactive.

Moreover, earlier digital health systems lacked integration with advanced technologies like machine learning. This restricts their ability to process and learn from large volumes of historical health data, identify hidden patterns, and improve over time. The absence of such intelligent models hinders scalability and limits the effectiveness of these systems in providing personalized, data-driven predictions. As a result, traditional methods and outdated digital tools fall short in delivering the speed, accuracy, and adaptability required for modern, patient-centered diabetes care.

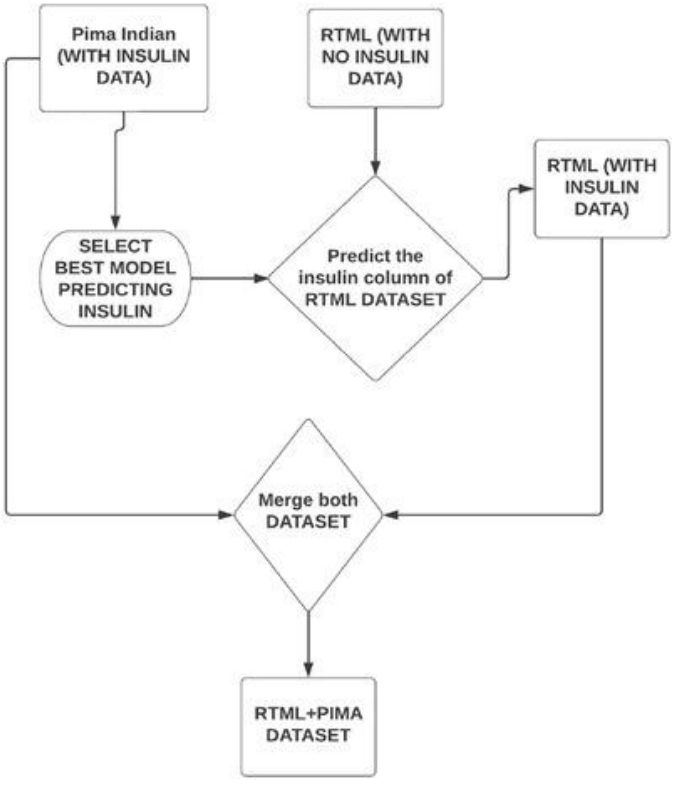

Fig -1: Work Flow

4.CONCLUSION

The DIABETES PREDICTION WEBSITE successfully addresses the need for an accessible and reliable tool to assist in early detection of diabetes using machine learning techniques. By leveraging trained models such as Logistic Regression or Random Forest and implementing a streamlined web interface via Flask, the system provides accurate predictions based on user input of health parameters like glucose level, BMI, insulin level, and blood pressure. The platform offers a user-friendly experience, enabling individuals—regardless of technical background—to input their health data and receive real-time feedback on their risk status.

The results are presented in a clear and concise format, making the tool suitable for educational purposes, self screening, and support in preliminary health assessment. The model delivers consistent and reliable predictions for typical use cases. However, for broader deployment, opportunities exist to further enhance the platform through real-time data integration (e.g., from health monitoring devices), expanding the model's training dataset for higher generalizability, and integrating advanced features such as personalized recommendations or visual health trend analysis.

In conclusion, the DIABETES PREDICTION WEBSITE represents a valuable step forward in the application of machine learning to healthcare. It demonstrates how artificial intelligence can contribute to the early detection and prevent the strategies which are done in real-time scenario, supporting individuals and healthcare providers alike. This project lays a solid foundation for further development towards a comprehensive digital health monitoring system.

ACKNOWLEDGEMENT

We express our deepest gratitude to our honorable Chairman, Joint Secretary, Principal, Head of the Department and our Project Mentor for their invaluable support, guidance, and cooperation throughout the research period, ensuring the successful completion of the project.

REFERENCES

1. Smith, J., & Brown, L. (2018). Diabetes Prediction Using Machine Learning Algorithms. *International Journal of Healthcare Informatics*, 5(2), 45–53. ISSN: 2456-5345.

2. Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., & Chouvarda, I. (2017). Machine Learning and Data Mining Methods in Diabetes Research. *Computational and Structural Biotechnology Journal*, 15, 104–116. ISSN: 2001-0370.

3. Zou, Q., Qu, K., Luo, Y., Yin, D., Ju, Y., & Tang, H. (2018). Predicting Diabetes Mellitus with Machine Learning Techniques. *Frontiers in Genetics*, 9, 515. ISBN: 1664-8021.

4. Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. ISSN: 978-1-4503-4232-2.

5. Akhtar, S. , et al.: Prevalence of diabetes and pre-diabetes in Bangladesh: A systematic review and meta-analysis. BMJ Open 10, e036086 (2020)

6. Mohan, N. , Jain, V. : Performance analysis of support vector machine in diabetes prediction. In: International Conference on Electronics, Communication and Aerospace Technology, pp. 1–3 (2020)

7. Mounika, V. , Neeli, D.S. , Sree, G.S. , Mourya, P. , Babu, M.A. : Prediction of type-2 diabetes using machine learning algorithms. In: International Conference on Artificial Intelligence and Smart Systems, pp. 127–131 (2021)

8. Kumari, S. , Kumar, D. , Mittal, M. : An ensemble approach for classification and prediction of diabetes mellitus using soft voting classifier. Int. J. Cognit. Comput. Eng. 2, 40–46 (2021)

9. Aurelien, G.: Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly Media, Inc., Sebastopol, CA

10. Chatrati, S.P., Hossain, G., Goyal, A., et al.: Smart home health monitoring system for predicting type 2 diabetes and hypertension. *J. King Saud Univ. Comput. Inf. Sci.* **34**(3), 862–870 (2020)

11. Cervantes, J., García-Lamont, F., Rodríguez, L., Lopez-Chau, A.: A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing* **408**, 189–215 (2020)

12. Hasan, M.K., Alam, M.A., Das, D., Hossain, E., Hasan, M.: Diabetes prediction using ensembling of different machine learning classifiers. *IEEE Access* **8**, 76516–76531, (2020)

13. Pranto, B., et al.: Evaluating machine learning methods for predicting diabetes among female patients in Bangladesh. *Information* **11**, 1–20 (2020)

14. Olisah, C.C., Smith, L., Smith, M.: Diabetes mellitus prediction and diagnosis from a data preprocessing and machine learning perspective. *Comput. Methods Programs Biomed.* **220**, 1–12 (2022)

15. Jackins, V., Vimal, S., Kaliappan, M., Lee, M.Y.: AI-based smart prediction of clinical disease using random forest classifier and Naive Bayes. *J. Supercomput.* **77**, 5198–5219 (2021)