Presentation on

DIABETES PREDICTION: WEB-BASE APP FOR EARLY DETECTION AND MONITORING

Presented By

Adarsh Lilhare

Arti Paraskar

Harshal Pawar

Shreyash Nangirwar

Bhupesh Wankhade

Guided By

Prof.L.S.Bhattad

Year 2024-25

Abstract

- Diabetes is a chronic disease affecting millions worldwide, characterized by elevated blood sugar levels. This project employs machine learning regression techniques to analyze the PIMA Indian Diabetes Dataset and enhances diagnostic accuracy by identifying key risk patterns. It enables early detection and proactive healthcare through a **web-based application**, allowing users to conveniently input data, access predictions, and view results via an interactive and user-friendly interface.

Introduction

- Diabetes is a growing global health concern with serious complications such as heart disease, kidney failure, and nerve damage. Traditional diagnostic methods are often slow and prone to errors, creating a need for automated solutions. This project uses machine learning regression techniques to analyze patient data and provide accurate, early-stage diabetes predictions. To increase accessibility and usability, the model is deployed through a web-based application that delivers real-time results via an interactive interface.

AIM & Objectives

The aim of our system is to develop a **web-based machine learning application** using regression techniques for early and accurate diabetes detection.

Our objective include:

- Explore the PIMA Indian Diabetes Dataset for patterns and key risk factors.

- Preprocess data to improve model quality and handle missing values.

- Apply regression algorithms to predict diabetes with precision.

- Integrate results into a user-friendly web interface.

- Evaluate performance using metrics like ROC-AUC, precision, and recall. *(Receiver Operating Characteristic- Area Under the Curve)*

Problem Statement

The current methods of diabetes diagnosis face several challenges:

- Conventional techniques are slow, costly, and heavily reliant on manual interpretation.

- Early detection is often delayed, increasing the risk of severe complications.

- Advanced machine learning-based models are needed to streamline diagnosis, improve accuracy, and support healthcare professionals in decision-making.

- This project addresses these challenges by deploying a real-time, web-accessible ML prediction tool.

Literature Survey

- **Regression for Diabetes Prediction:** Studies have shown the effectiveness of regression models in assessing the relationships between key health variables and diabetes risk.

- **Feature Importance Analysis:** Research highlights the significance of features such as glucose levels, BMI, and insulin in improving diagnostic precision.

- **AI in Early Diagnosis:** Machine learning and regression-based techniques demonstrate transformative potential in healthcare solutions for disease management.

- **Web-based:** Web-based applications are increasingly used in healthcare for patient monitoring and support.

Idea

- The core idea is to build a **web-based diabetes prediction system** that allows users to input data and receive predictions in real-time
- This involves integrating ML regression models into a responsive and secure web application with features like:

    - User authentication
    - OTP verification
    - Chatbot support
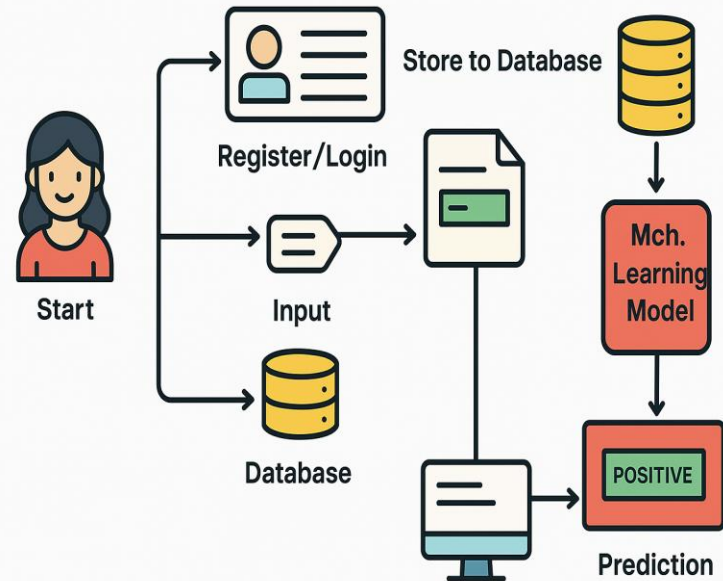
Methodology

- **Data Collection**: PIMA Diabetes Dataset
- **Data Preprocessing**: Handling missing values, normalization
- **Evaluation**: ROC-AUC, accuracy, precision, recall
- **Deployment**: Flask-based web app, SQLite for history, MySQL for user data
- **Regression Models Applied:** Logistic Regression as the core model.
- **Evaluation Metrics:** Accuracy, precision, recall, F1-score, and ROC-AUC for model comparison.
- **Result Integration:** Develop a weighted voting ensemble approach for final prediction.

Screen shots/ Results
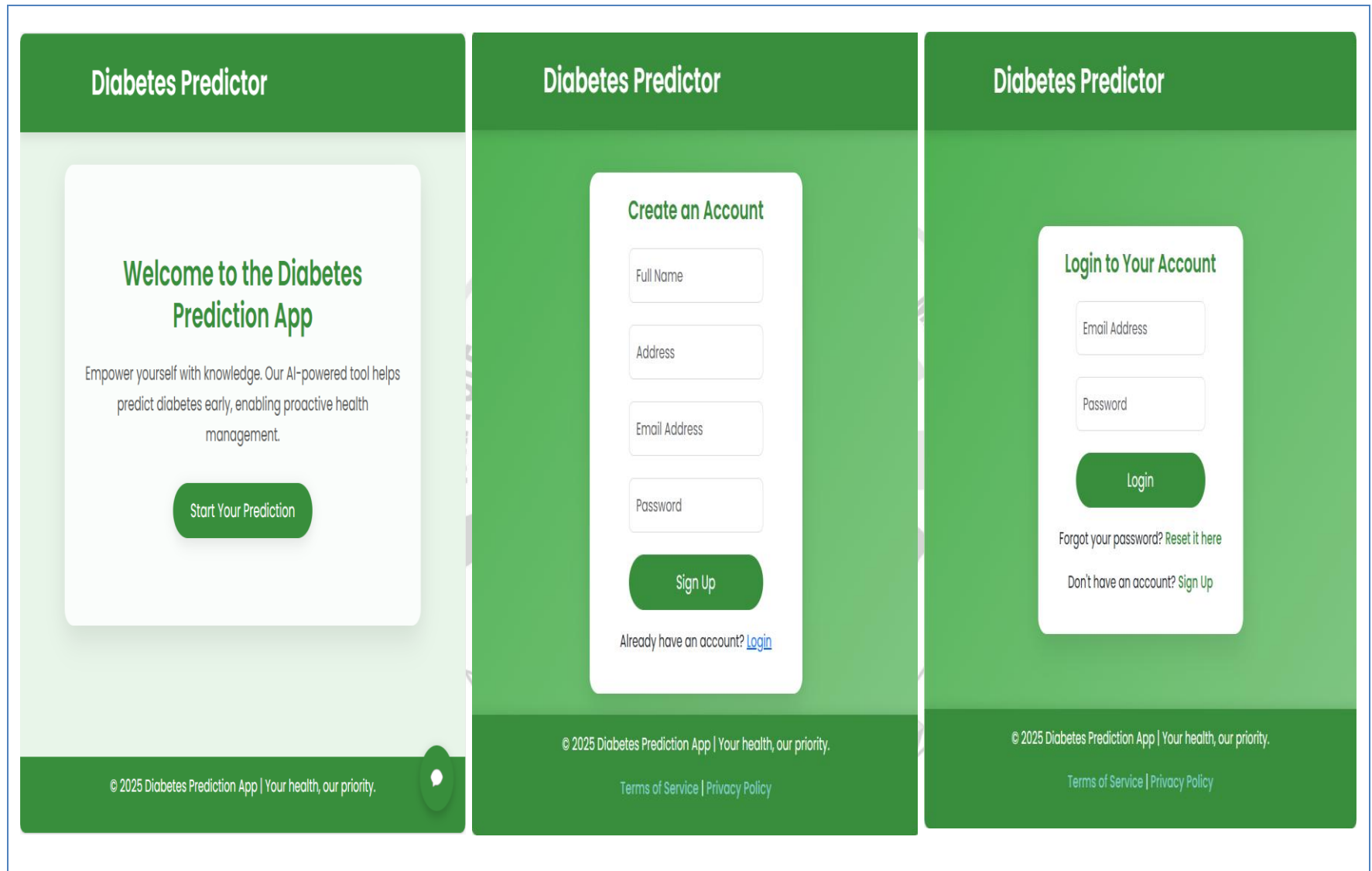
Screen shots/ Results

Screen shots/ Results
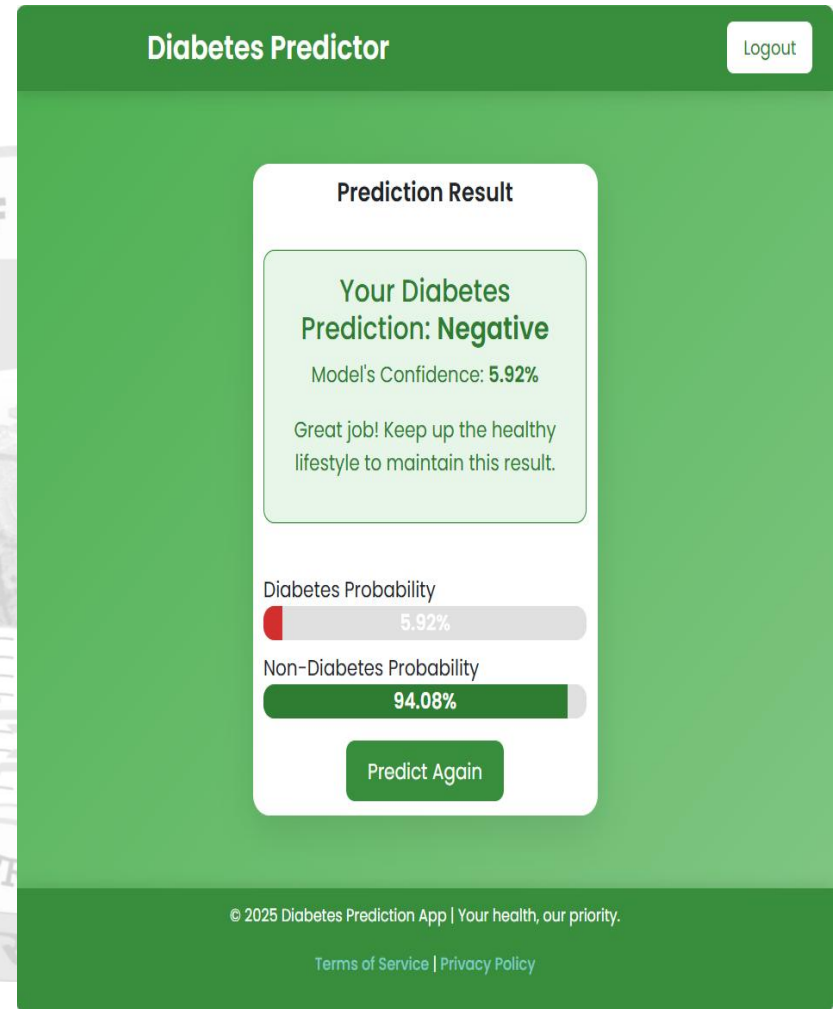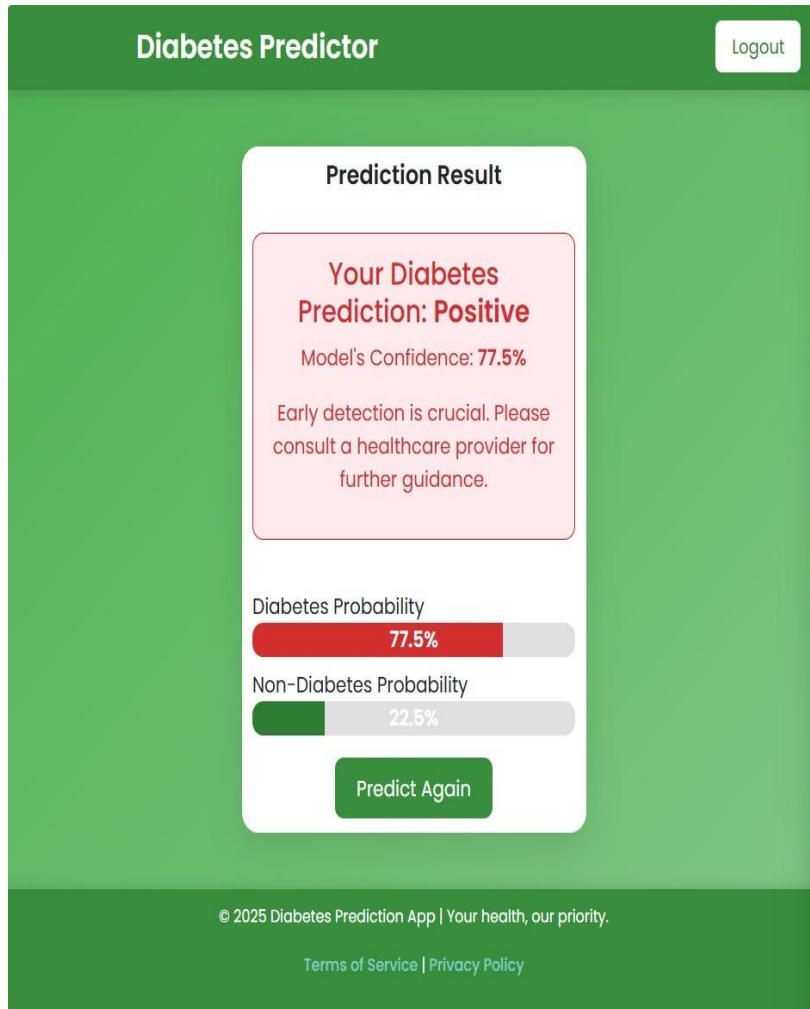
Result Analysis

- Regression models identified strong correlations between patient data and diabetes risk, providing reliable predictions.
- Logistic Regression achieved the highest accuracy and ROC-AUC score of 0.950, indicating excellent classification performance.
- The final ensemble-based approach demonstrated high accuracy and reliability for early diabetes prediction.
- Data preprocessing, including feature scaling and handling missing values, significantly improved model reliability.
- The regression approach highlights simplicity, interpretability, and efficiency for diabetes diagnosis.
- Web deployment provides a real-time, scalable solution.

Opportunities and Challenges

Opportunities:
- Regression-based models facilitate early and accurate diabetes prediction.
- Predictive models empower healthcare professionals with accurate data insights.
- Reduces healthcare costs by streamlining diagnosis.

Challenges:
- Managing imbalanced datasets and incomplete patient information.
- Addressing concerns around data privacy and ethical use of sensitive health data.
- Computational complexity in real-time prediction models.

Conclusion

This project highlights the potential of regression-based machine learning for accurate diabetes prediction. The integration with a web-based application enhances accessibility and user convenience. It enables early diagnosis and assists healthcare professionals in making informed decisions. The system minimizes human error and offers cost-effective, real-time predictions. With further development, it can be expanded to include more features and datasets for improved accuracy.

References

Journal Article:

[1] R. T. Prasad, S. B. Nagajyothi, and G. Harika, "Diabetes Prediction Using Ensembling of Different Machine Learning Classifiers," *International Journal of Scientific Research in Engineering and Management (IJSREM)*, vol. 06, no. 06, pp. 1–5, Jun. 2022.

[2] A. Kadam and S. Zade, "Diabetes Prediction Using Machine Learning," *International Research Journal of Engineering and Technology (IRJET)*, vol. 08, no. 06, pp. 3526–3530, Jun. 2021.

[3] D. Shinde, P. Patil, A. Varpe, P. Ghadge, and R. Gawade, "Diabetes Prediction using Machine Learning," *International Journal of Engineering Research & Technology (IJERT)*, vol. 10, no. 12, pp. 1–4, Dec. 2021.

References

Website:

[1] Pima Indians Diabetes Dataset: https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database/ (UCI ML Repository)

[2]International Diabetes Federation. (2021). IDF Diabetes Atlas, 10th edition. https://diabetesatlas.org/

[3]Machine Learning Mastery. (2020). A Tour of Machine Learning Algorithms. https://machinelearningmastery.com/atour-of-machine-learning-algorithms/