GlucoSense: AI-Powered Diabetes Detection for Early Intervention

Presented by: Varun Myaka

Glucosense leverages AI to identify early warning signs of diabetes, empowering individuals and healthcare providers to make proactive decisions for better health outcomes.

Problem Statement

- Diabetes is a growing global health concern, affecting millions globally, impacting quality of life and leading to serious complications.
- Early detection is crucial for effective management and prevention of long-term health issues.
- This project aims to develop an AI-powered solution to detect diabetes at an early stage by analyzing key patient data. By leveraging machine learning algorithms, the model predicts diabetes risk, empowering healthcare providers to take timely action and enhance patient care.

Data Collection

- Diabetes dataset has 520 entries, contains information relevant to diagnosing diabetes.

- Key features in the dataset likely include demographic details (e.g., Age, Gender) and symptom indicators (e.g., Polyuria, Polydipsia, visual blurring).

- The target variable is class, which indicates whether a person is Positive or Negative for diabetes.

Feature Engineering Approach

1

Feature Selection

Identifying the most relevant features that contribute significantly to predicting diabetes.

2

Feature Extraction

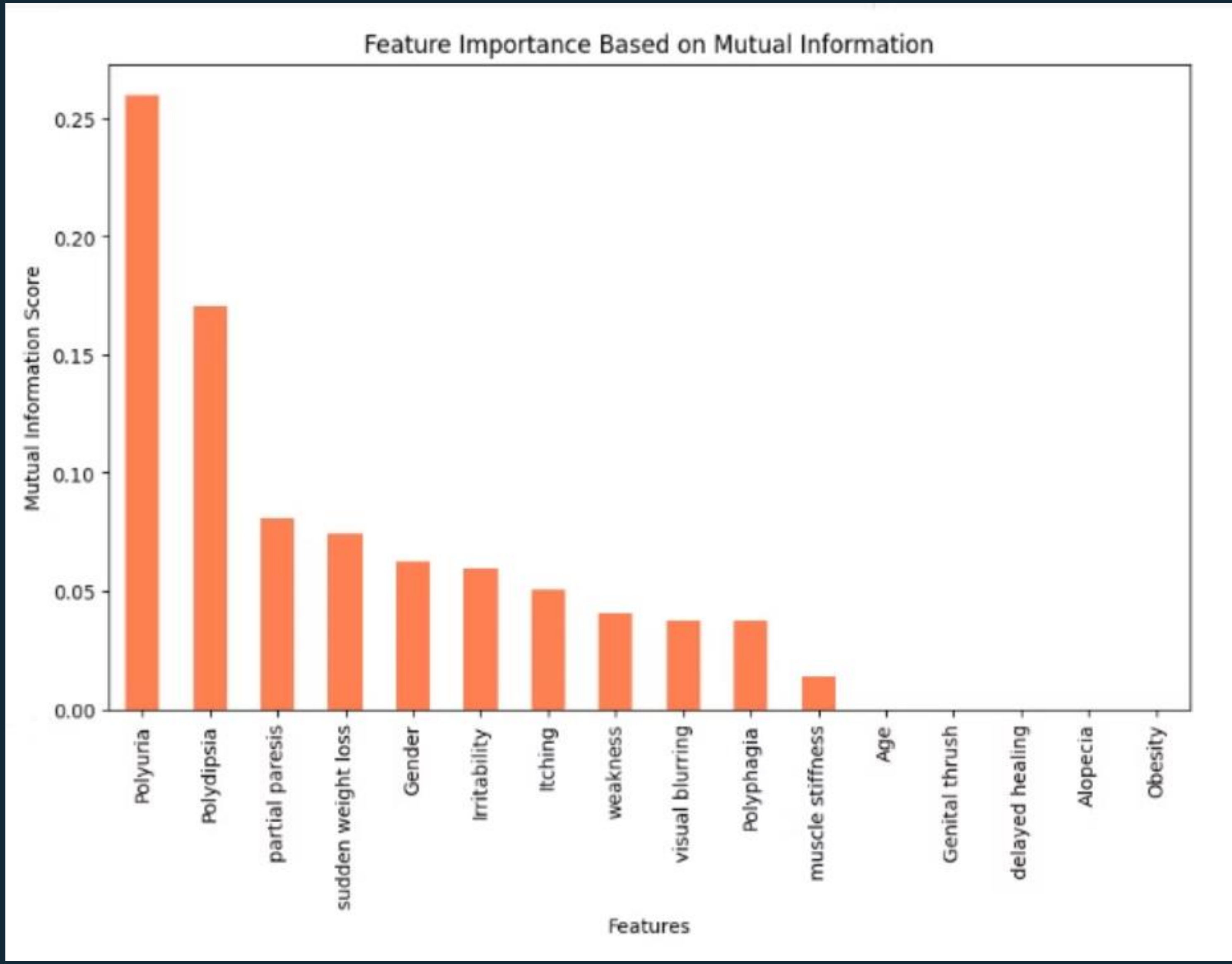
Creating new features from existing data to enhance model performance.

3

Feature Transformation

Transforming existing features into a more suitable format for the model.

Feature Importance



Data Preprocessing Methods

Handling Missing Values:

- Identified and replaced missing data using the **mean, median**, or **mode** where applicable.

Encoding Categorical Variables:

- Converted categorical variables (e.g., "Yes/No") into binary values (0/1) for machine learning compatibility.

Feature Scaling:

- Standardized numerical features using **StandardScaler** to ensure all features are on a comparable scale.

Balancing the Dataset:

- Addressed class imbalance using techniques like oversampling (if applied) to improve model performance on minority classes.

Outcomes of Data Preprocessing

Clean and Structured Dataset:

- Dataset prepared with **250 entries** and all features encoded for model training.

Significant Features Identified:

- Glucose levels, BMI, and age were identified as key predictors of diabetes risk through correlation analysis.

Improved Model Readiness:

- Preprocessed data ensured better convergence and higher accuracy during model training.

Eliminated Redundancy and Noise:

- Removal of outliers and irrelevant features reduced noise, leading to more robust models.

Models Evaluated

1

Logistic Regression

A statistical model used to predict the probability of an event occurring.

2

Random Forest

An ensemble learning method that combines multiple decision trees to improve accuracy.

3

Decision Tree

A simple, tree-like model used for both classification and regression.

4

Gradient Boosting

An ensemble method that combines multiple decision trees sequentially, improving accuracy.

5

Support Vector Machines (SVM)

A supervised learning model used for classification and regression tasks.

6

Naive Bayes

A probabilistic classifier based on Bayes' theorem.

7

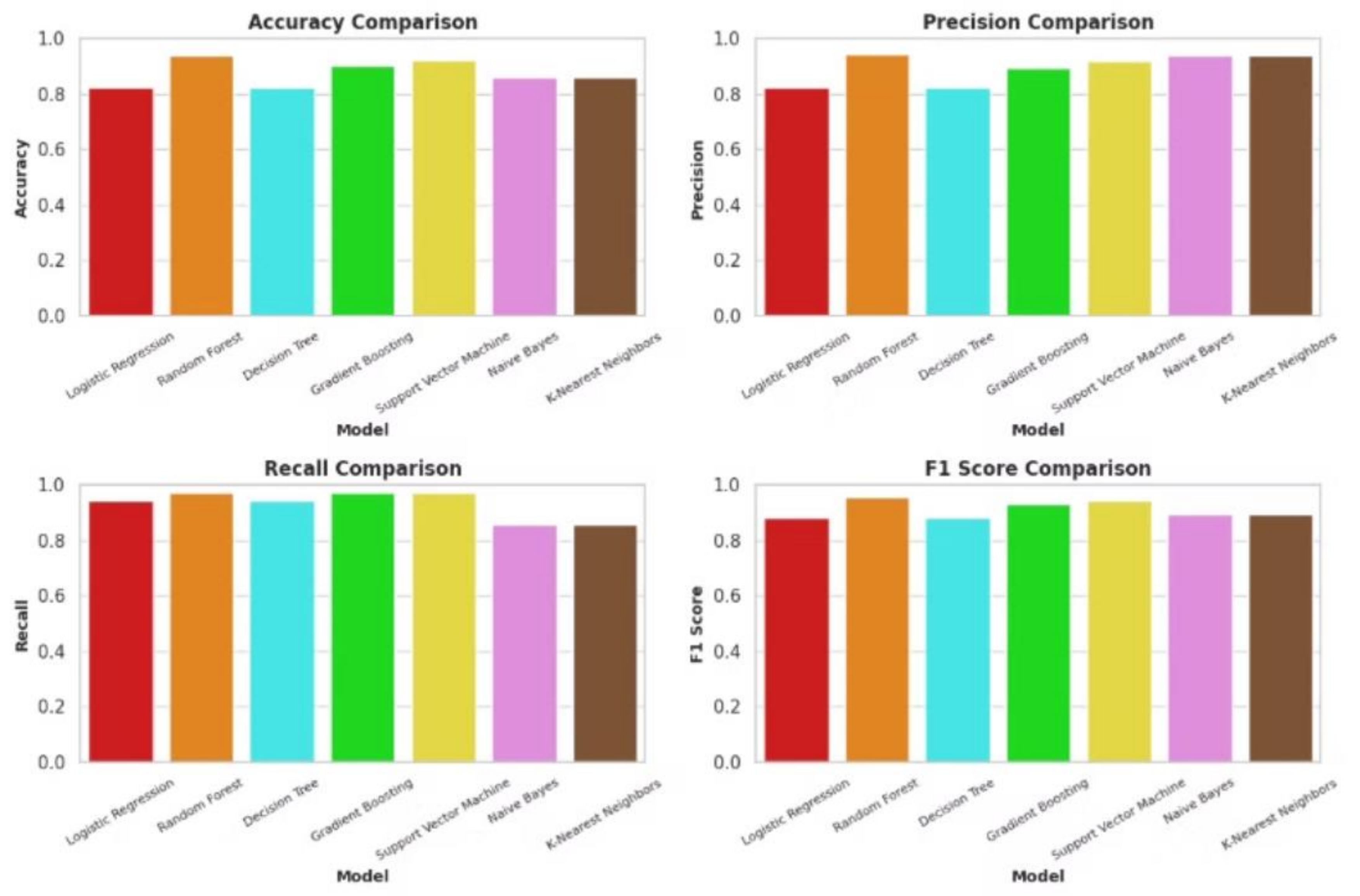
KNN

A simple algorithm that classifies data points based on the majority class among its nearest neighbors.

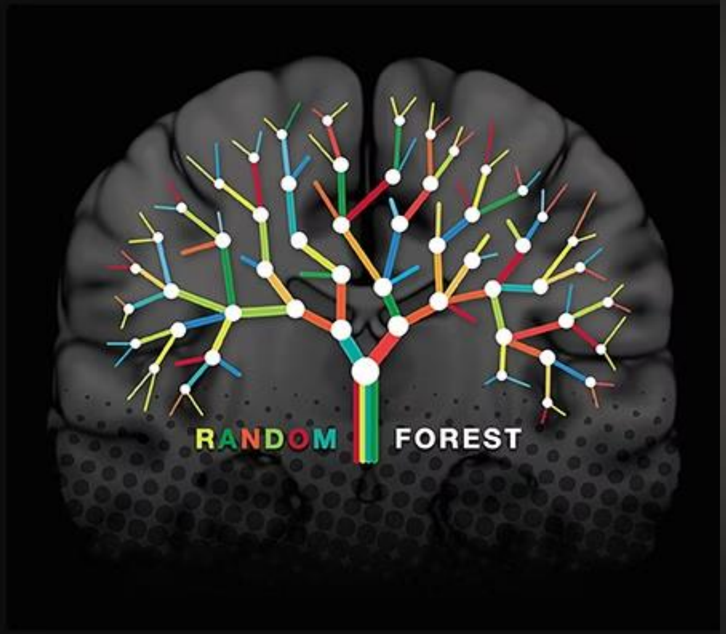
Model Building Methodology

- **Data Splitting:** Divided the dataset into training (80%) and testing (20%) subsets to ensure unbiased evaluation of the models.
- **Feature Selection:** Identified critical features influencing diabetes prediction, such as Glucose Level, BMI, and Family History.
- **Imbalanced Dataset Handling:** Addressed the class imbalance by utilizing F1-score as the evaluation metric to prioritize precision and recall.
- **Evaluation Metrics:** Compared model performance based on F1-score to identify the most suitable algorithm for diabetes detection.
- **Best Model Identified:** Random Forest was selected as the optimal model for its balanced performance across metrics.

Model Performance Comparision



Made with Gamma



Performance Metrics Evaluated

(Random Forest)

94%

Accuracy

The proportion of correctly classified instances.

94%

Precision

The ability of the model to correctly identify positive cases.

97%

Recall

The ability of the model to identify all positive cases.

95%

F1-Score

A harmonic mean of precision and recall, balancing both metrics.

Performance Metrics and Evaluation

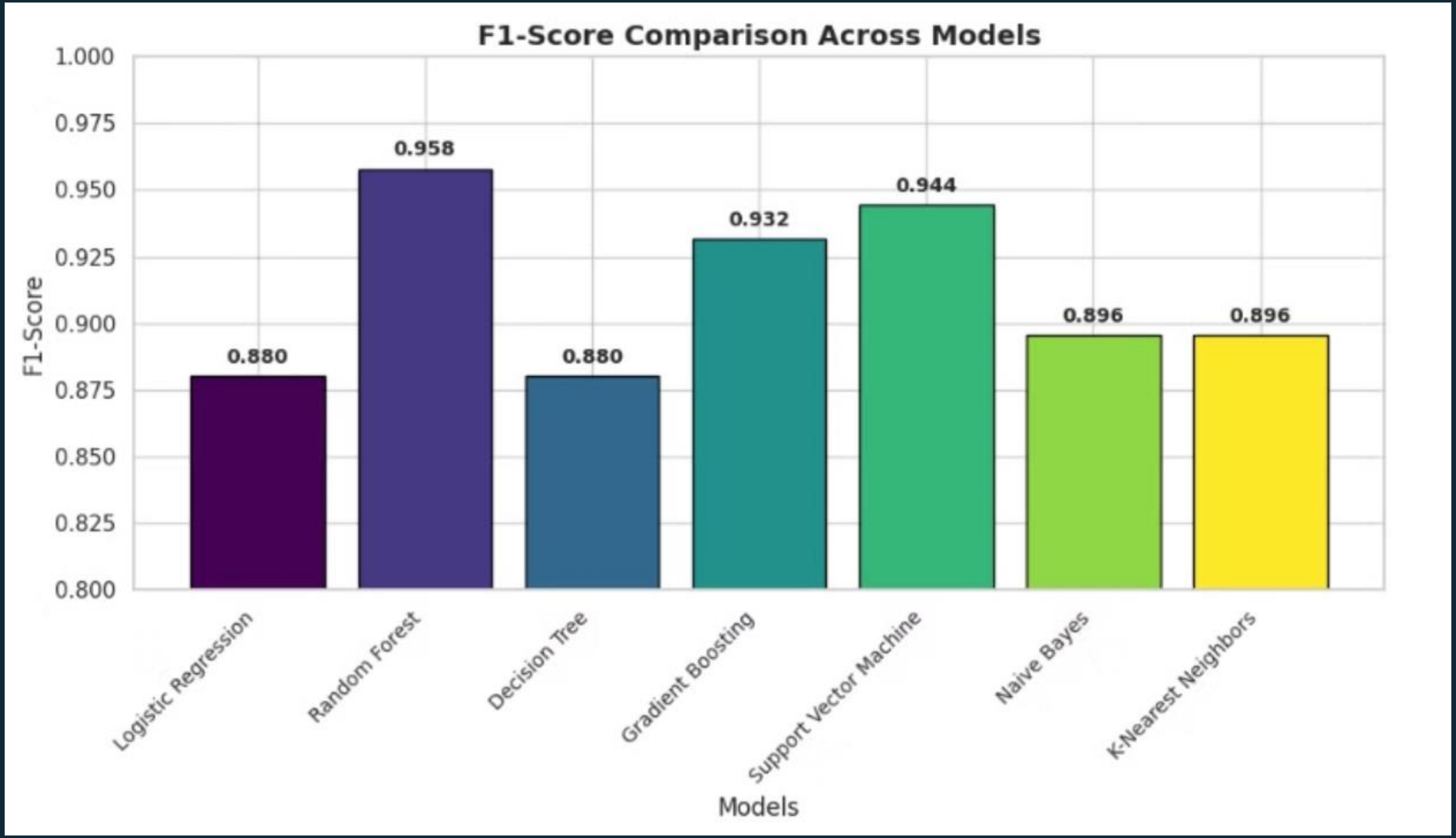
Evaluation Metric Used:

- **F1-Score:** Selected due to the imbalanced nature of the dataset, balancing precision and recall effectively.

Key Observations:

- Random Forest consistently outperformed other models in terms of F1-score, making it the most reliable choice for diabetes prediction in this project.
- Other models showed varying degrees of trade-offs between precision and recall.

F1-Score Comparison Across Models



Recommendations

Adopt Random Forest Model:

- With the highest F1-score, the Random Forest model is the most suitable for predicting diabetes in this dataset.
- Its ability to handle imbalanced data and provide robust predictions makes it ideal for deployment.

Future Improvements:

- Collect more diverse and balanced data.
- Include additional features such as lifestyle habits, diet specifics, and genetic predispositions for better predictions.
- Explore deep learning for further accuracy.

Deployment Recommendations:

- Integrate the model into healthcare systems for screening.
- Develop a mobile or desktop application for usability.

Recommendations for Early Intervention

Early Detection and Treatment:

- GlucoSense can aid in early detection, enabling timely interventions and preventing complications.

Lifestyle Modifications:

- The model's insights can guide individuals in adopting healthier habits, like diet and exercise, to manage blood glucose levels.

Personalized Care Plans:

- The system can generate personalized care plans based on individual risk factors, facilitating proactive management.

Regular Monitoring:

- GlucoSense encourages frequent monitoring of blood glucose levels, allowing for timely adjustments to treatment plans.

Thank You for Your Time

I appreciate your interest.