

# Diabetes Prediction using SVM

A comprehensive overview of utilizing **machine learning** to predict diabetes effectively using Support Vector Machine classifiers.



Pooja Divekar 23101C0049

Kirtana Singh 23101C0065

Pradnya Khanse 23101C0051



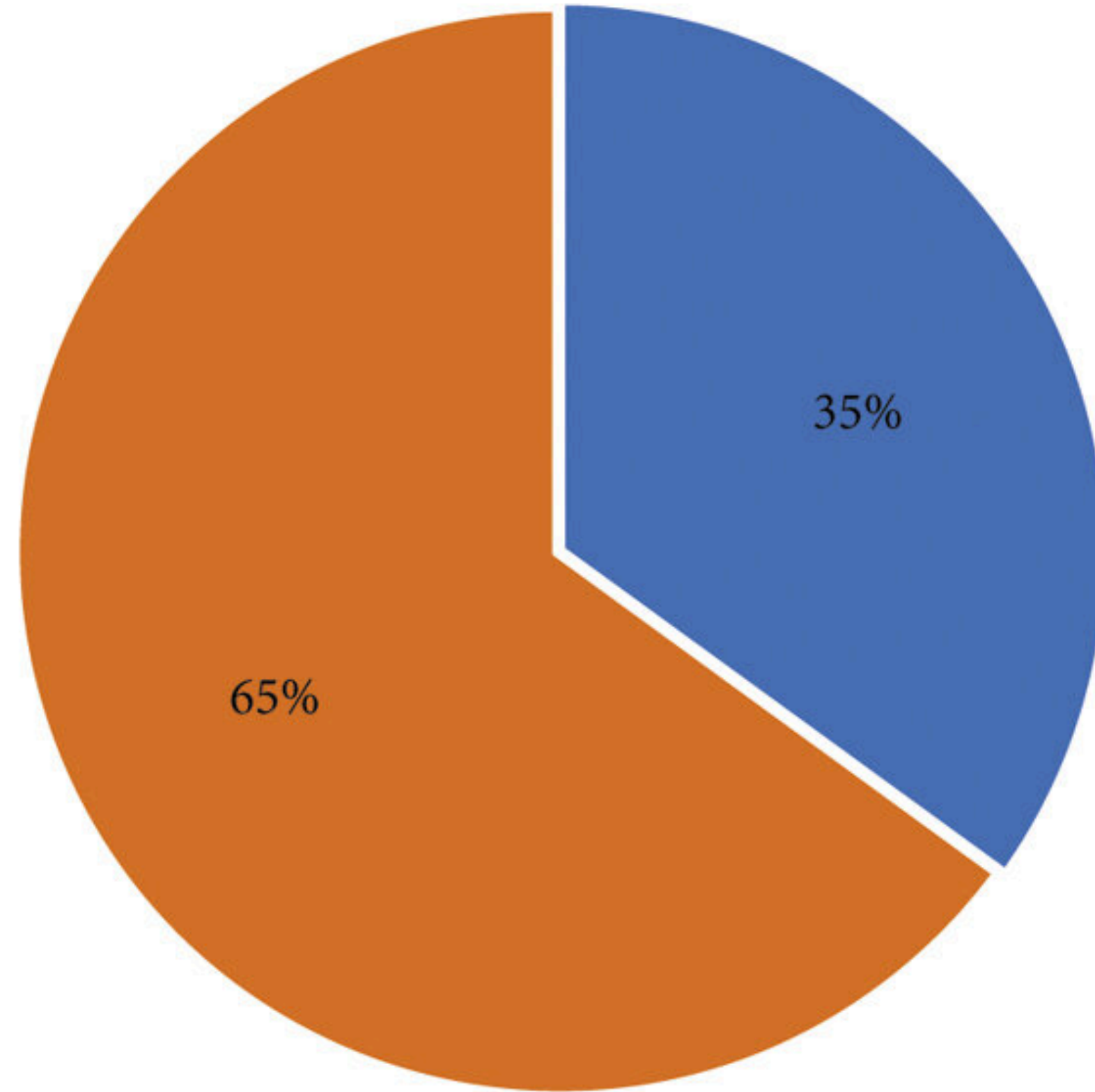


# Early Detection Matters

- Diabetes is a chronic disease caused by high blood sugar levels.
- Early detection helps prevent complications like heart disease, kidney failure, etc.
- Machine Learning allows for data-driven health prediction.
- This project applies Support Vector Machine (SVM) for diabetes classification.



Percentage of diabetes among Pima  
Indians participants



- Diabetic - Outcome (1)
- Not Diabetic - Outcome (0)

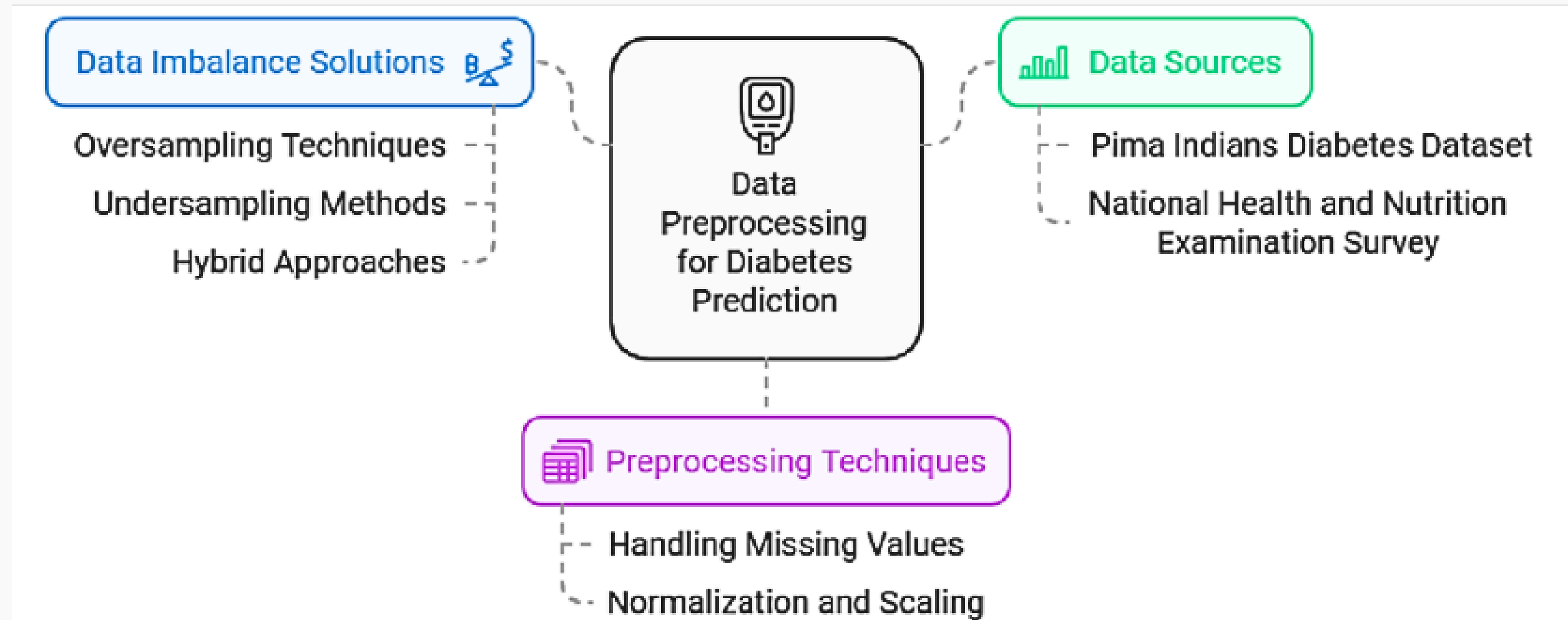
# Dataset Overview

- Dataset Source: PIMA Indian Diabetes Dataset
- Instances: 767 samples
- Attributes (9 total):
  - a. Pregnancies
  - b. Glucose
  - c. BloodPressure
  - d. SkinThickness
  - e. Insulin
  - f. BMI
  - g. DiabetesPedigreeFunction
  - h. Age
  - i. Outcome (1 = Diabetic, 0 = Non-Diabetic)

Observation: No missing values, but several zero entries indicate missing physiological data.



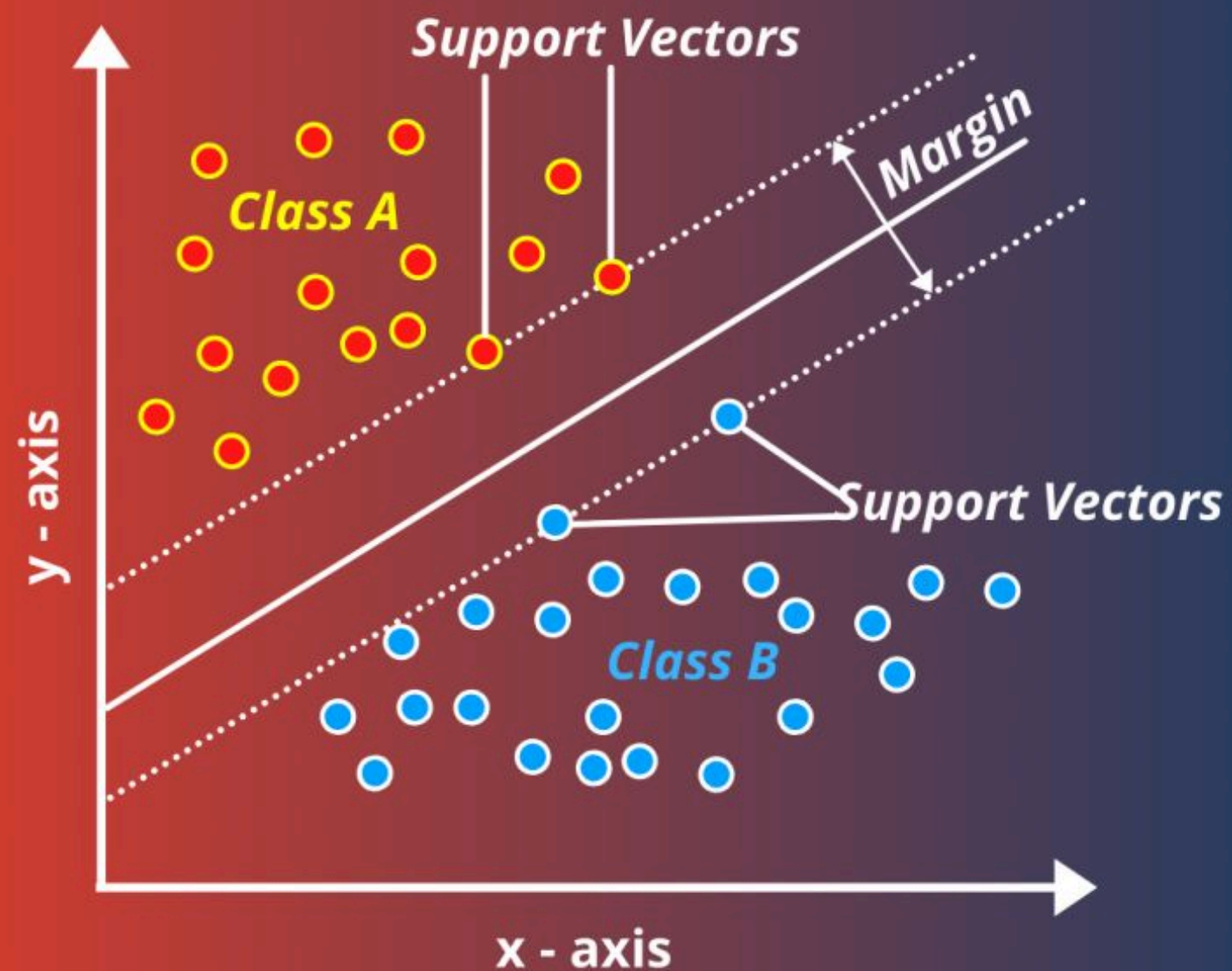




# Data Preprocessing Steps

Data preprocessing is crucial for building an effective model. We renamed unnamed columns, applied **StandardScaler** for normalization, and partitioned the dataset into training and testing sets, ensuring an 80%-20% split. Furthermore, we analyzed correlations among features to uncover valuable relationships prior to modeling.

# SUPPORT VECTOR MACHINE



# Support Vector Machine (SVM)

The **Support Vector Machine** algorithm is a powerful tool for classification tasks, especially in medical diagnostics. It works by finding the optimal hyperplane that separates different classes, in this case, diabetic and non-diabetic patients. Its robustness and effectiveness make it ideal for handling complex biological data.

Model trained using:

```
model = SVC(kernel="linear", probability=True)
```



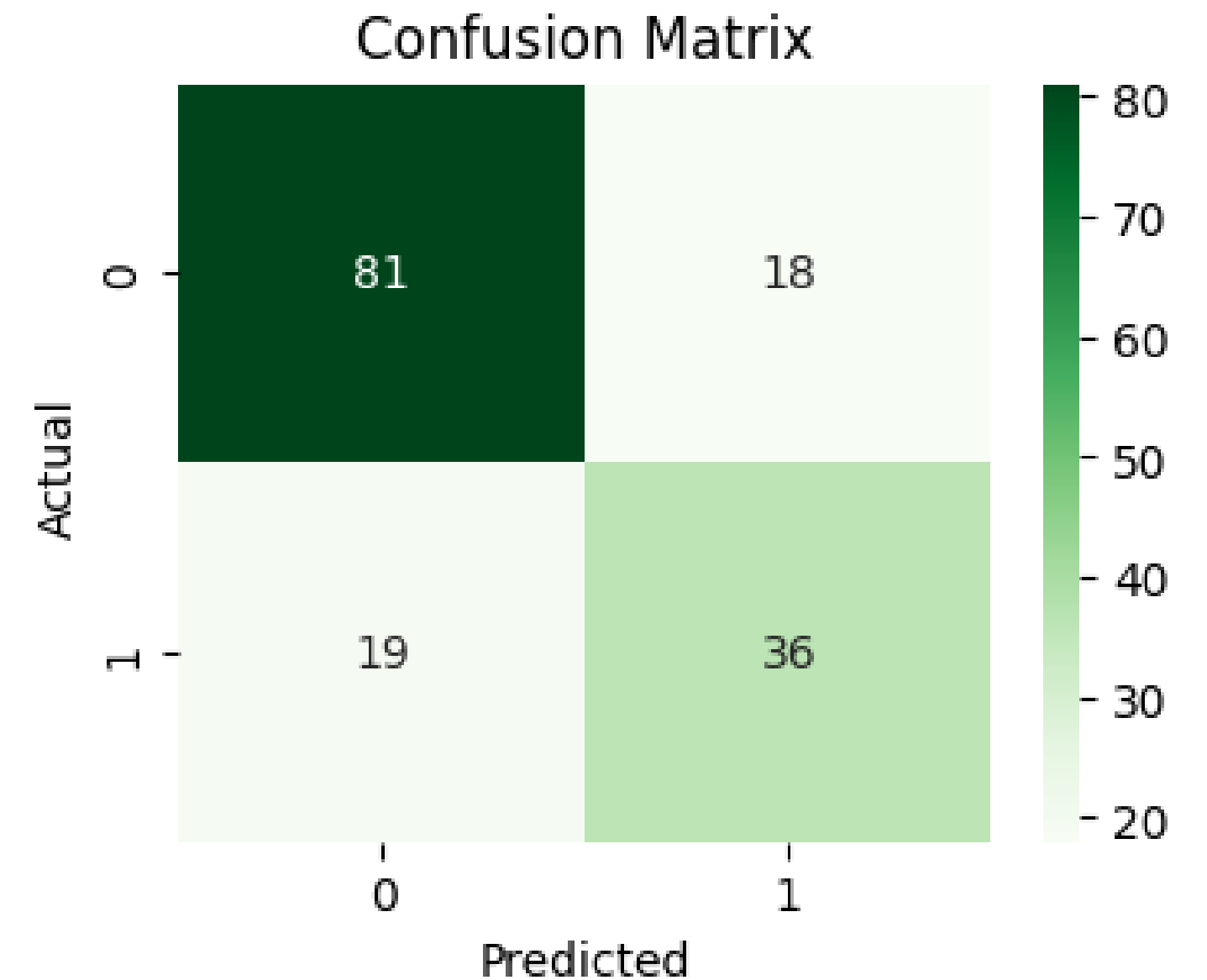
# Model Evaluation

- Accuracy: ~ (replace with your model's output, e.g. 78–82%)
- Metrics Used:
  - Confusion Matrix
  - Accuracy Score
- Example Visualization:
  - Heatmap of confusion matrix
  - Glucose vs Age scatter plot

Interpretation: Glucose and BMI are strong predictors of diabetes.

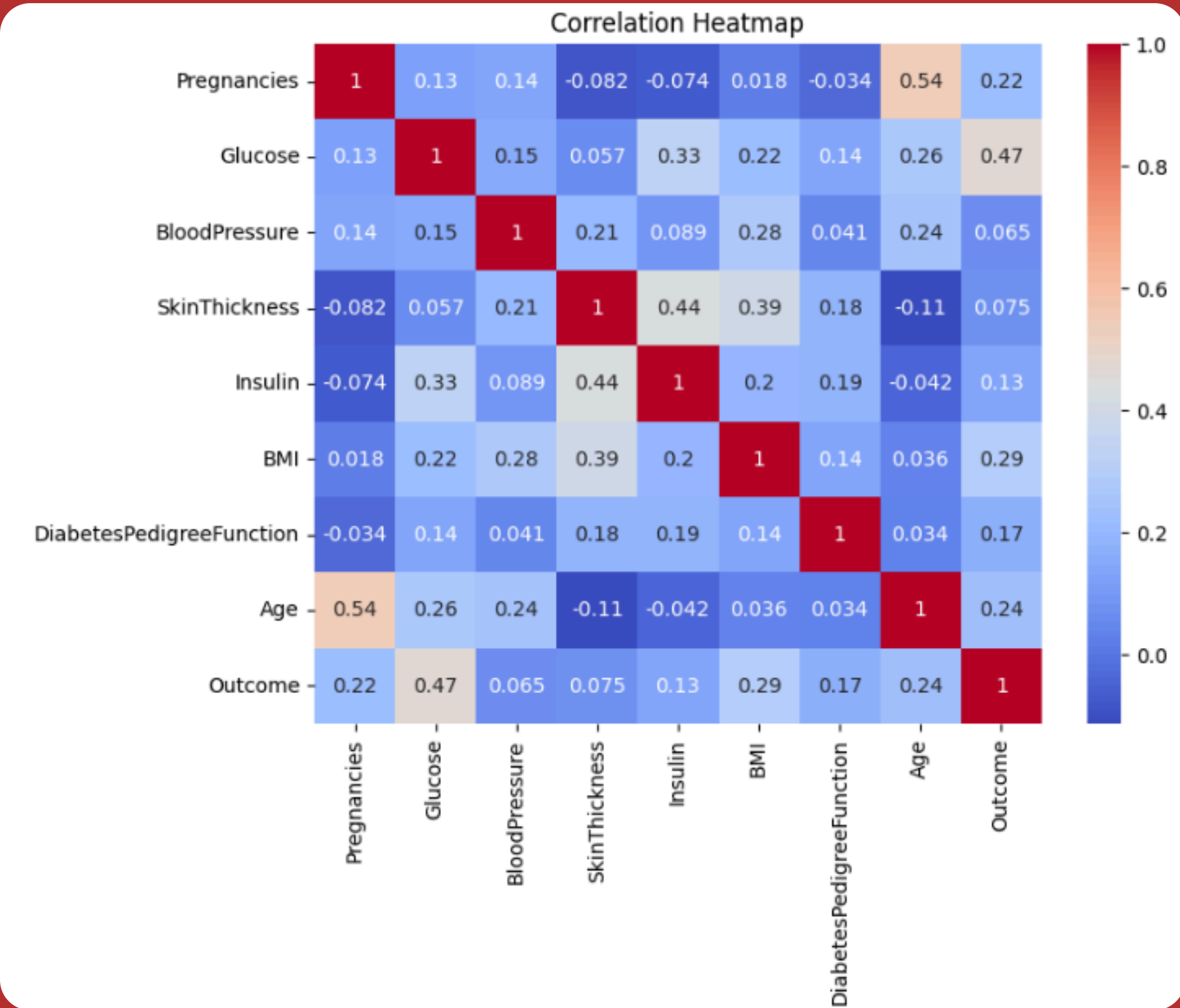
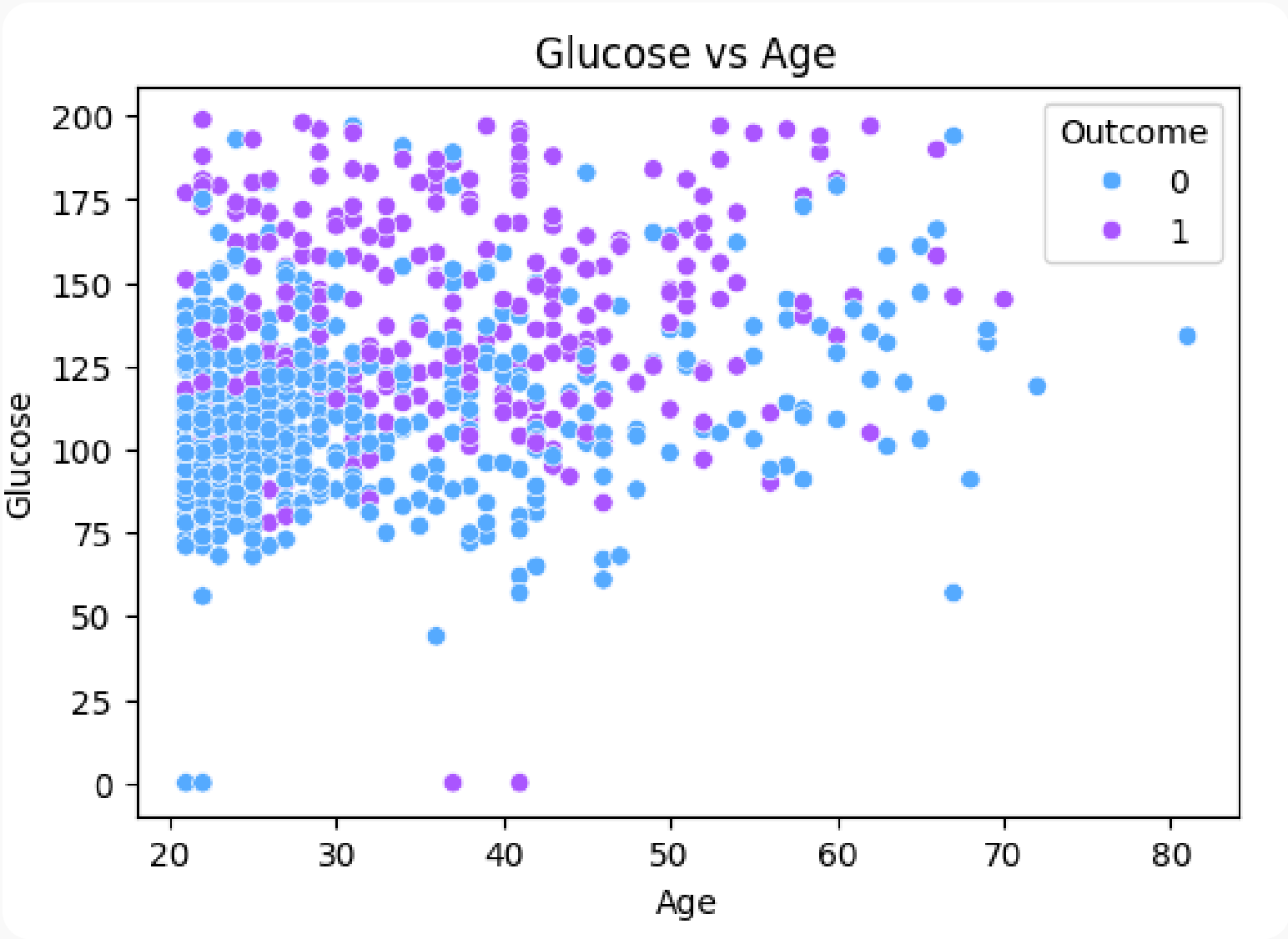


✓ Model trained successfully with accuracy: 75.97%



# Data Visualization

This chart illustrates the **relationships among features**, highlighting the correlation between Glucose and Age. It reveals that older individuals with higher glucose levels are more likely to be diabetic, emphasizing critical predictors.





## Diabetes Prediction using SVM

Gender:	<div>Female</div>
Pregnancies:	<div><div></div></div> 5
Glucose:	<div><div></div></div> 102
BP:	<div><div></div></div> 73
Skin Thk:	<div><div></div></div> 52
Insulin:	<div><div></div></div> 340
BMI:	<div><div></div></div> 30.60
DPF:	<div><div></div></div> 0.56
Age:	<div><div></div></div> 43

Predict

# Interactive Prediction System

- Built with ipywidgets for real-time user input.
- Users enter their values:
- Glucose, BMI, Insulin, Age, etc.
- System displays:
-  “Not Diabetic” or  “Likely Diabetic”
- Confidence percentage and model accuracy
- Smart UI hides “Pregnancy” slider when gender = Male.



# Conclusion & Future Work

## Conclusion:

- SVM classifier effectively predicts diabetes with strong accuracy.
- Visualization improves understanding of key health parameters.

## Future Enhancements:

- Add advanced models (Random Forest, Neural Networks).
- Deploy as a web or mobile app for easy use.
- Integrate medical datasets for broader applicability.

