

PIMA INDIANS DIABETES



Introduction

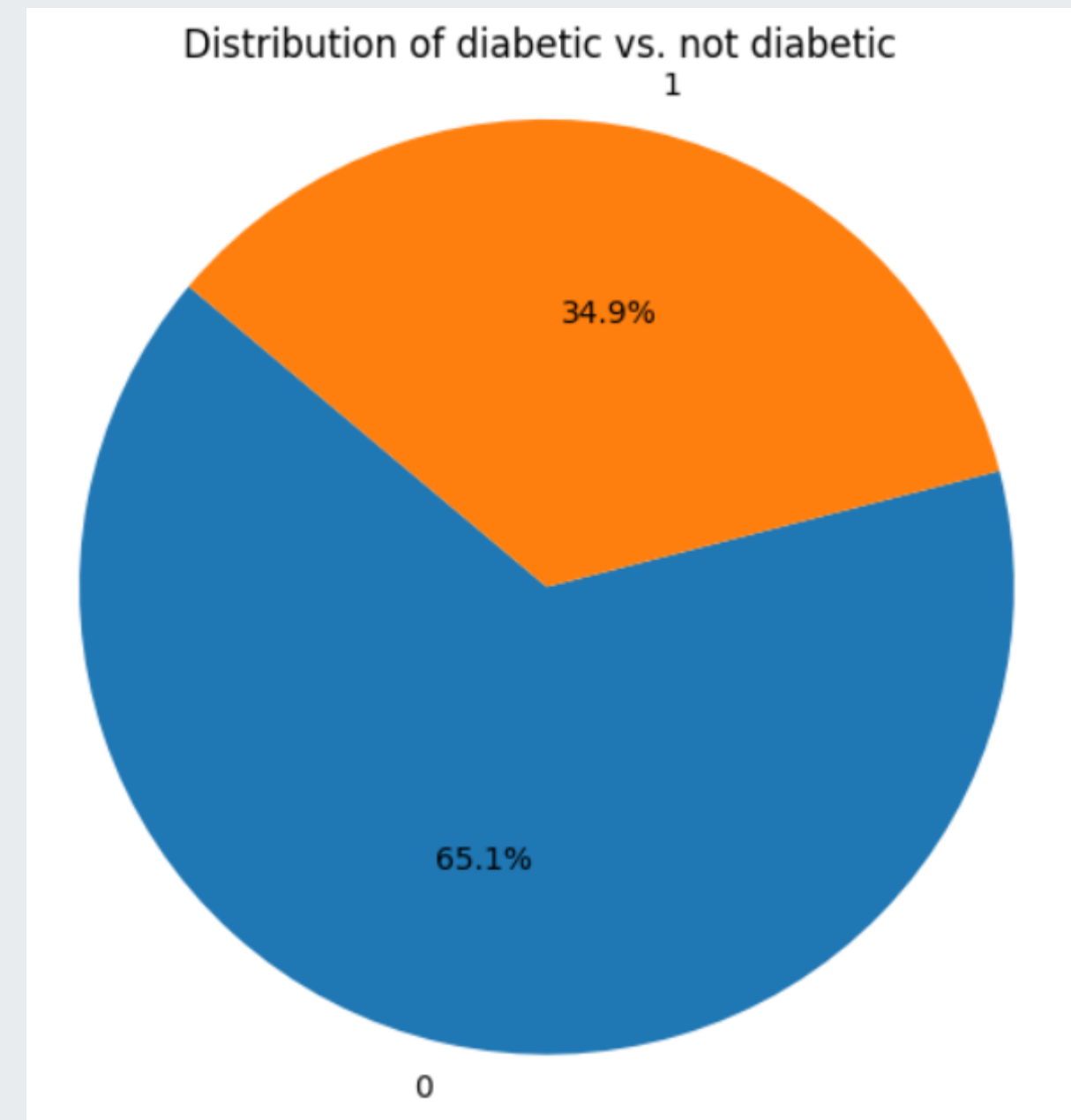
The dataset focuses on predicting diabetes in Pima Indian women and encounters challenges due to missing data and the complexity of forecasting health conditions with limited attributes. Researchers explore various machine learning algorithms like Random Forest, K-Nearest Neighbors, Naive Bayes, Neural Network with Dense Layers, Linear Discriminant Analysis, Principal Component Analysis, Decision Tree, and Bayesian Belief Network. Their primary aim is early diabetes detection, evaluating models using accuracy metrics to select suitable models for practical real-world predictions.



Methodology

by using outcome column and applying pie chart on it.

It represents the percentage of diabetic people which is 34.9% and not diabetic people which is 65.1%



Gaussian Naive Bayes for
classification :
with Training Accuracy: 0.75
Testing Accuracy: 0.75
so, it's balanced model

Bayesian Belief Network
structure :
[('Pregnancies', 'Outcome'),
('Glucose', 'Outcome')]

Decision Tree :
Training Accuracy: 0.75 Testing
Accuracy: 0.72

PCA :
Training Accuracy: 0.75
Testing Accuracy: 0.67





LDA :
Training Accuracy: 0.75 Testing
Accuracy: 0.71

Neural Network :
Accuracy 0.71

KNN :
Accuracy with Manhattan
distance: 0.73 Accuracy with
Chebyshev distance: 0.70

K-fold cross-validation :
Training Accuracy: 0.65 Testing
Accuracy: 0.63

Results



Naive Bayes

Accuracy of Gaussian Naive Bayes: 0.75

Decision Tree

Accuracy of Decision Tree: 0.72. Model is balanced fit.

PCA

Accuracy with PCA: 0.67.
Model is balanced fit.

Bayesian Belief Network

Bayesian Network structure: [('Pregnancies', 'Outcome'), ('Glucose', 'Outcome')]

Neural Network

Accuracy 0.75. model is balanced fit.

LDA

Accuracy with LDA: 0.71. Model is balanced fit.

conclusion

The Pima Indians Diabetes dataset is crucial for predicting diabetes in Pima Indian women.

Models used :

- Naive Bayes
- Bayesian Belief Network
- Decision Tree
- KNN
- Neural Network
- PCA
- LDA

