Heart Disease
Prediction

Canva

Group Members:
- Shruti Kumari-2105495
- Sanu Verma-21051509
- Siddhanth Sargam-21051516
- K Tanvi Ananya-21051523
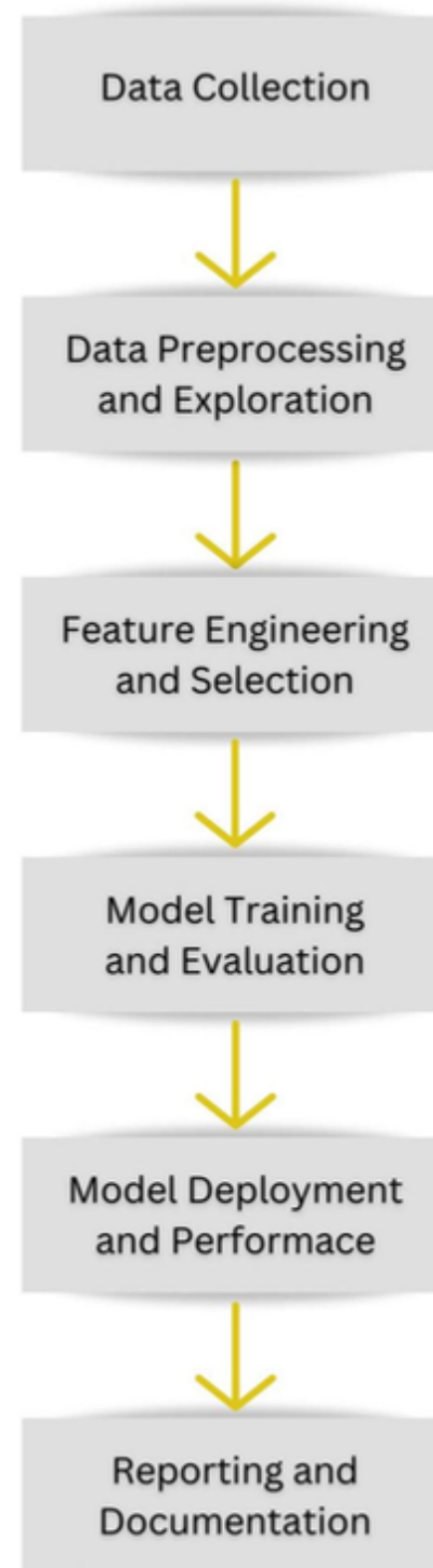- Utkrist Jaiswal-21051526

Objective

To develop a predictive model that can accurately assess the risk of an individual developing heart disease based on various factors such as demographic information, medical history, lifestyle choices, and diagnostic test results. The ultimate goal is to aid healthcare professionals in identifying individuals who are at high risk of developing heart disease so that appropriate interventions, preventive measures, and treatments can be implemented to mitigate the risk and improve patient outcomes.

Introduction

Heart disease presents a substantial global health challenge, jeopardizing both individual well-being and healthcare systems. Conventional diagnostic methods frequently encounter difficulties in effectively analyzing intricate patient data, resulting in limitations in disease detection. To address this, our project harnesses machine learning algorithms to revolutionize heart disease diagnosis. Through the utilization of extensive datasets and advanced predictive modeling techniques, our objective is to create a sophisticated diagnostic tool capable of accurately pinpointing subtle indicators of heart disease. This pioneering approach holds promise in enhancing patient outcomes and optimizing the allocation of healthcare resources.

Implementation

Data Collection

↓

Data Preprocessing and Exploration

↓

Feature Engineering and Selection

↓

Model Training and Evaluation

↓

Model Deployment and Performace

↓

Reporting and Documentation

In this segment, we elaborate on the execution of the heart disease prediction initiative, encompassing the methodology, testing and validation strategy, analysis of results, and quality control.

Data Collection

Data Preprocessing:

Model Development:

Model Evaluation:

Model Selection

```
Train Result:
========================================
Accuracy Score: 100.00%

Classification Report:   Precision Score: 100.00%
                         Recall Score: 100.00%
                         F1 score: 100.00%

Confusion Matrix:
 [[68  0]
 [ 0 83]]

Test Result:
========================================
Accuracy Score: 84.87%

Classification Report:   Precision Score: 87.34%
                         Recall Score: 84.15%
                         F1 score: 85.71%

Confusion Matrix:
 [[60 10]
 [13 69]]
```

```
Train Result:
========================================
Accuracy Score: 100.00%

Classification Report:   Precision Score: 100.00%
                         Recall Score: 100.00%
                         F1 score: 100.00%

Confusion Matrix:
 [[68  0]
 [ 0 83]]

Test Result:
========================================
Accuracy Score: 76.32%

Classification Report:   Precision Score: 80.26%
                         Recall Score: 74.39%
                         F1 score: 77.22%

Confusion Matrix:
 [[55 15]
 [21 61]]
```

```
Train Result:
========================================
Accuracy Score: 100.00%

Classification Report:   Precision Score: 100.00%
                         Recall Score: 100.00%
                         F1 score: 100.00%

Confusion Matrix:
 [[68  0]
 [ 0 83]]

Test Result:
========================================
Accuracy Score: 73.68%

Classification Report:   Precision Score: 76.25%
                         Recall Score: 74.39%
                         F1 score: 75.31%

Confusion Matrix:
 [[51 19]
 [21 61]]
```

Result

Conclusion

We analyzed a diverse heart health dataset, ensuring data integrity and exploring correlations through EDA. After preprocessing, we trained Random Forest, Gradient Boosting, and Decision Tree classifiers. Random Forest achieved 100% accuracy on training and 84.87% on testing, outperforming others. Our findings highlight ML's potential in predicting heart disease, with Random Forest showing promising results for clinical use. Further refinement and data collection could enhance diagnostic tools.
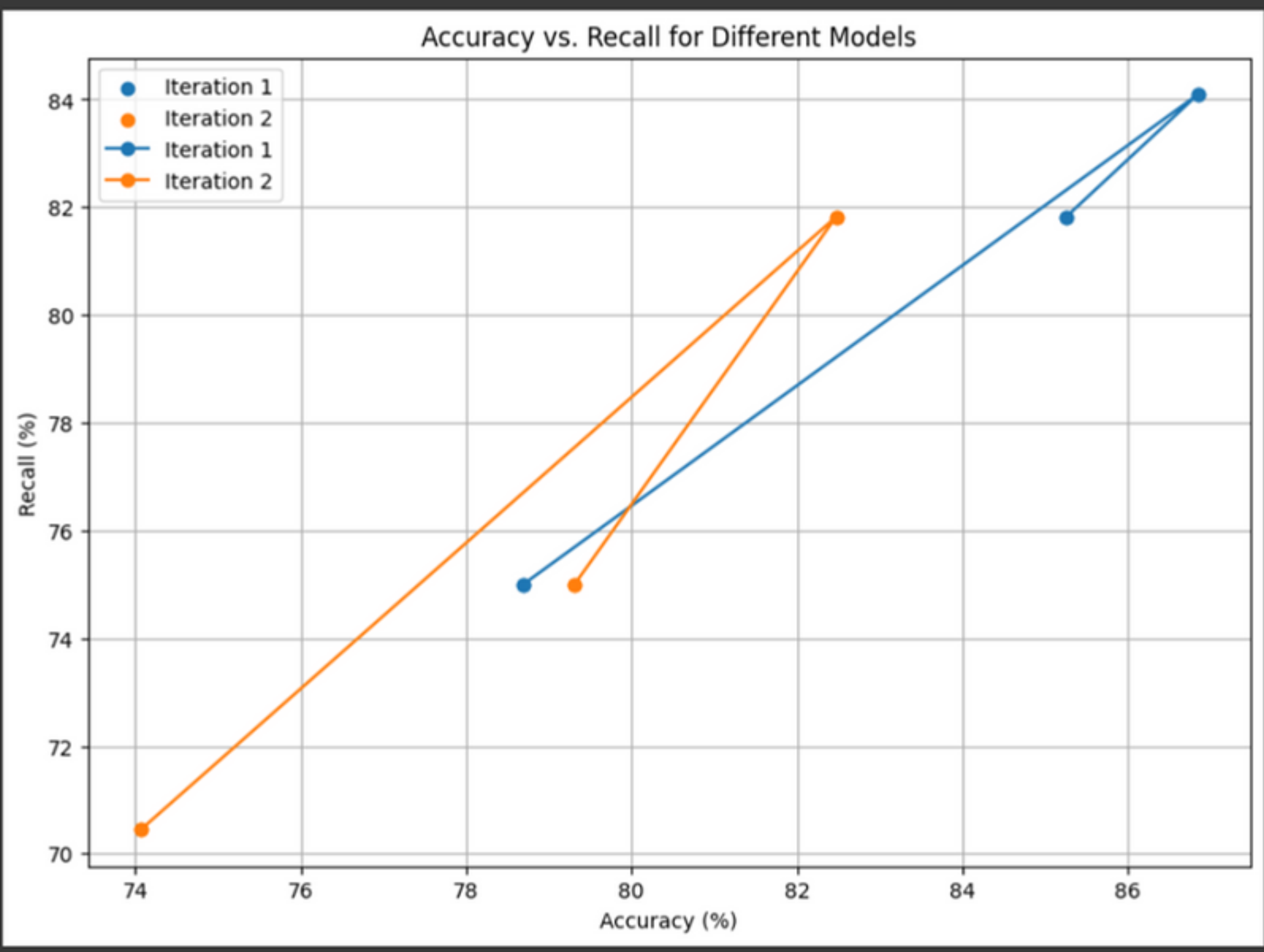
Future Scope

Future enhancements for this project include refining feature engineering, optimizing model parameters, exploring ensemble methods, addressing class imbalances, integrating models into clinical systems for real-time monitoring, improving interpretability, utilizing longitudinal studies, validating with external datasets, incorporating diverse data modalities, and establishing mechanisms for ongoing model refinement. These advances aim to enhance predictive analytics in cardiovascular health for improved preventive strategies and patient care.

Accuracy vs. Recall for Different Models

Thank You