CS-19411 PYTHON PROGRAMMING FOR MACHINE LEARNING

## 

## REPORT ON PPML PROJECT

Heart Disease Prediction Using Machine Learning

JANARTHANAN M | 220901032 | III YR EEE A

DEPARTMENT OF ELECTRICAL AND ELECTRONICS ENGINEERING

RAJALAKSHI ENGINEERING COLLEGE, THANDALAM

## Abstract

The **Heart Disease Prediction Using Machine Learning** project aims to develop a predictive model to identify the likelihood of heart disease in patients based on their medical history and physiological parameters, such as age, cholesterol levels, blood pressure, and heart rate. Leveraging the **UCI Heart Disease dataset**, this project explores various machine learning algorithms, including **Logistic Regression** and **Random Forest**, to build a reliable classifier.

Data preprocessing techniques were applied to ensure data quality, such as feature scaling and handling missing values. The **Random Forest model** was selected for its higher accuracy and ability to handle complex non-linear relationships, while **Logistic Regression** served as a baseline due to its simplicity and interpretability. Both models were trained and evaluated using standard metrics like **accuracy**, the **confusion matrix**, and a **classification report**, with Random Forest outperforming Logistic Regression. Additionally, **feature importance analysis** provided insights into the most influential factors in predicting heart disease.

This project demonstrates the practical application of machine learning in healthcare, offering a tool to assist in early detection and potentially reduce the risk of severe outcomes. Through this process, we gained a deeper understanding of key machine learning concepts, such as **model evaluation**, **feature selection**, and the trade-off between model complexity and interpretability.

I’m including the live demonstration of my project in the following pages in the Implementation Section !

## Literature Survey

Prior research highlights the growing use of machine learning for heart disease detection.

Comparative studies show Random Forest's robustness in handling non-linear data, while Logistic Regression provides a simpler, interpretable baseline.

Existing approaches often lack in-depth feature analysis, which this project addresses.

DATASET AND PREPROCESSING

**Dataset Source:** UCI Heart Disease dataset.

**Features:**

* Patient information: Age, Gender.
* Medical data: Cholesterol levels, Blood pressure, Resting heart rate.

**Preprocessing Techniques:**

* Handling missing values through imputation.
* Normalizing features using Min-Max Scaling.

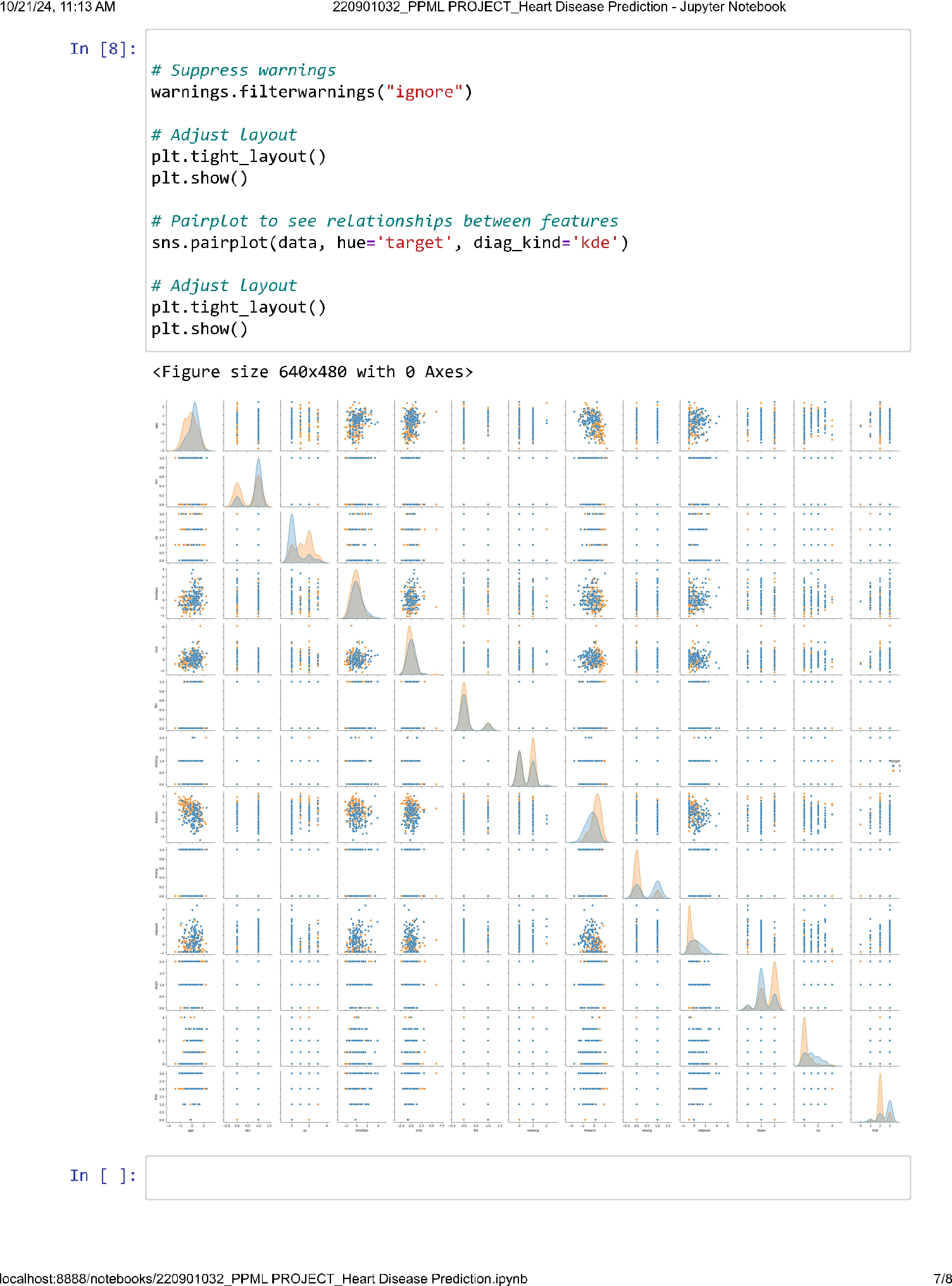
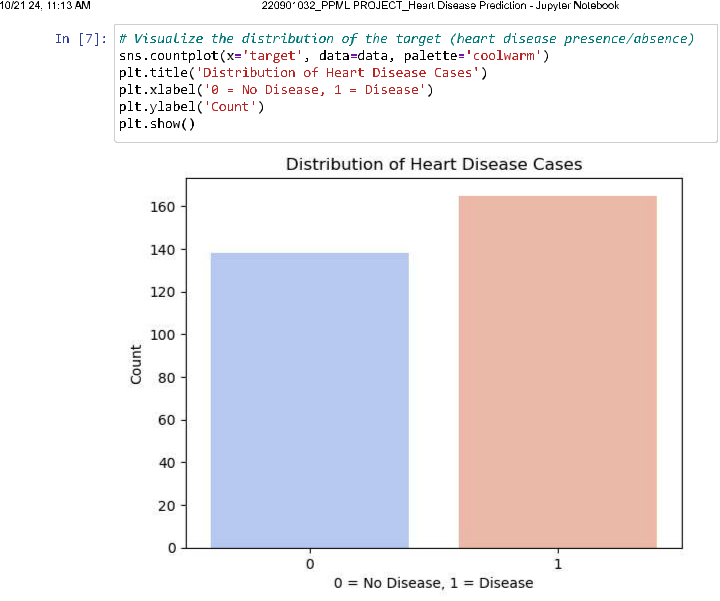
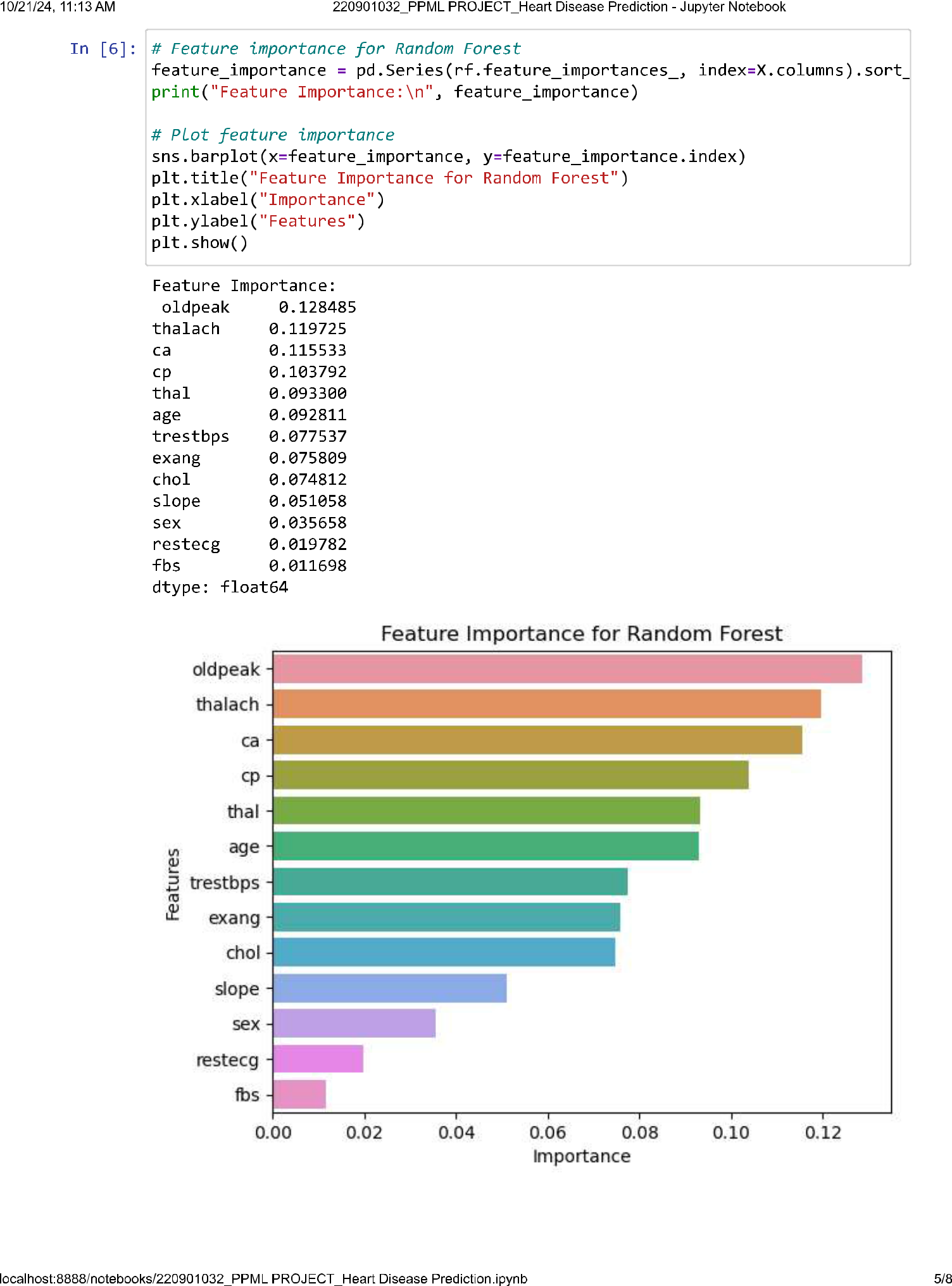
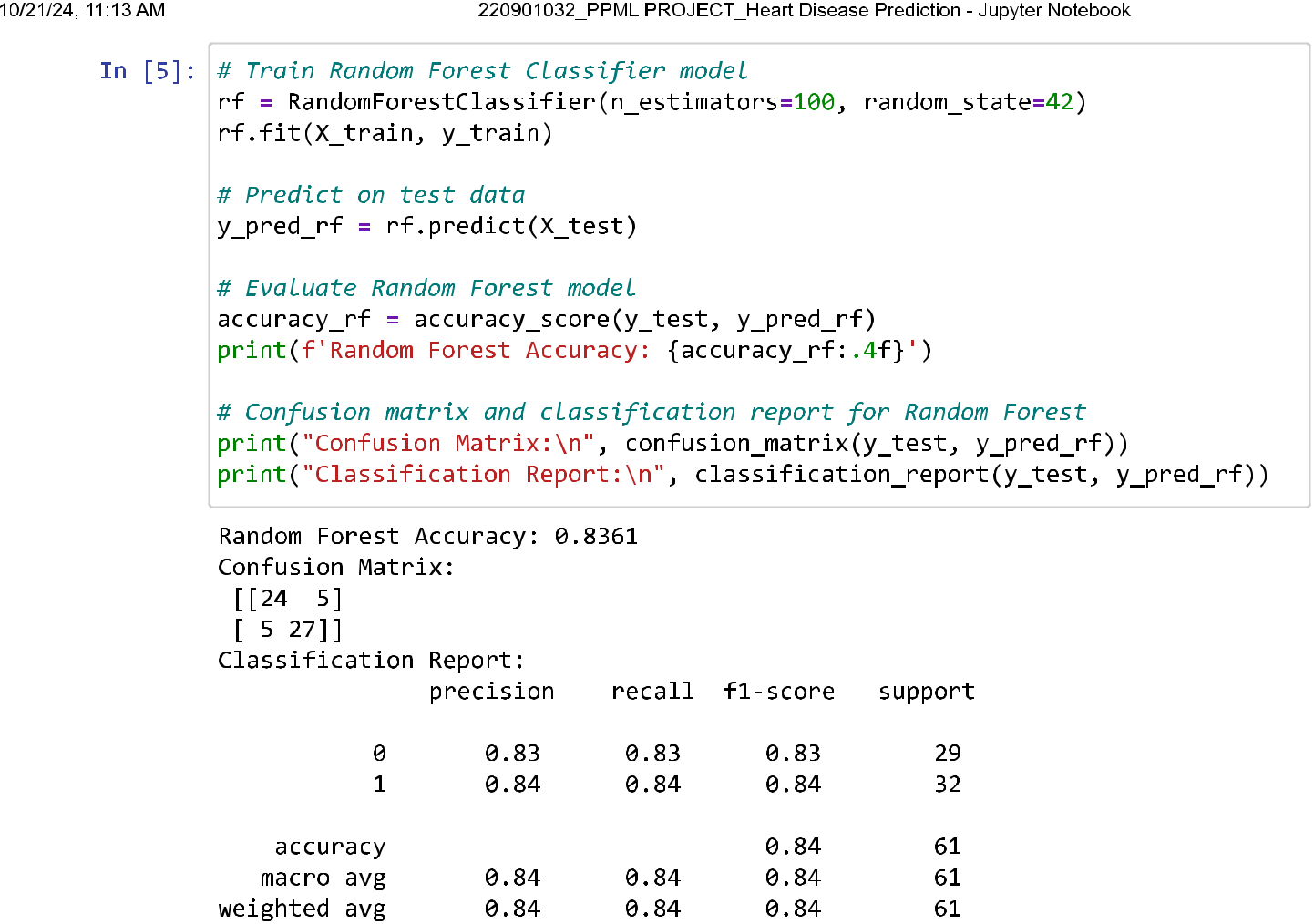
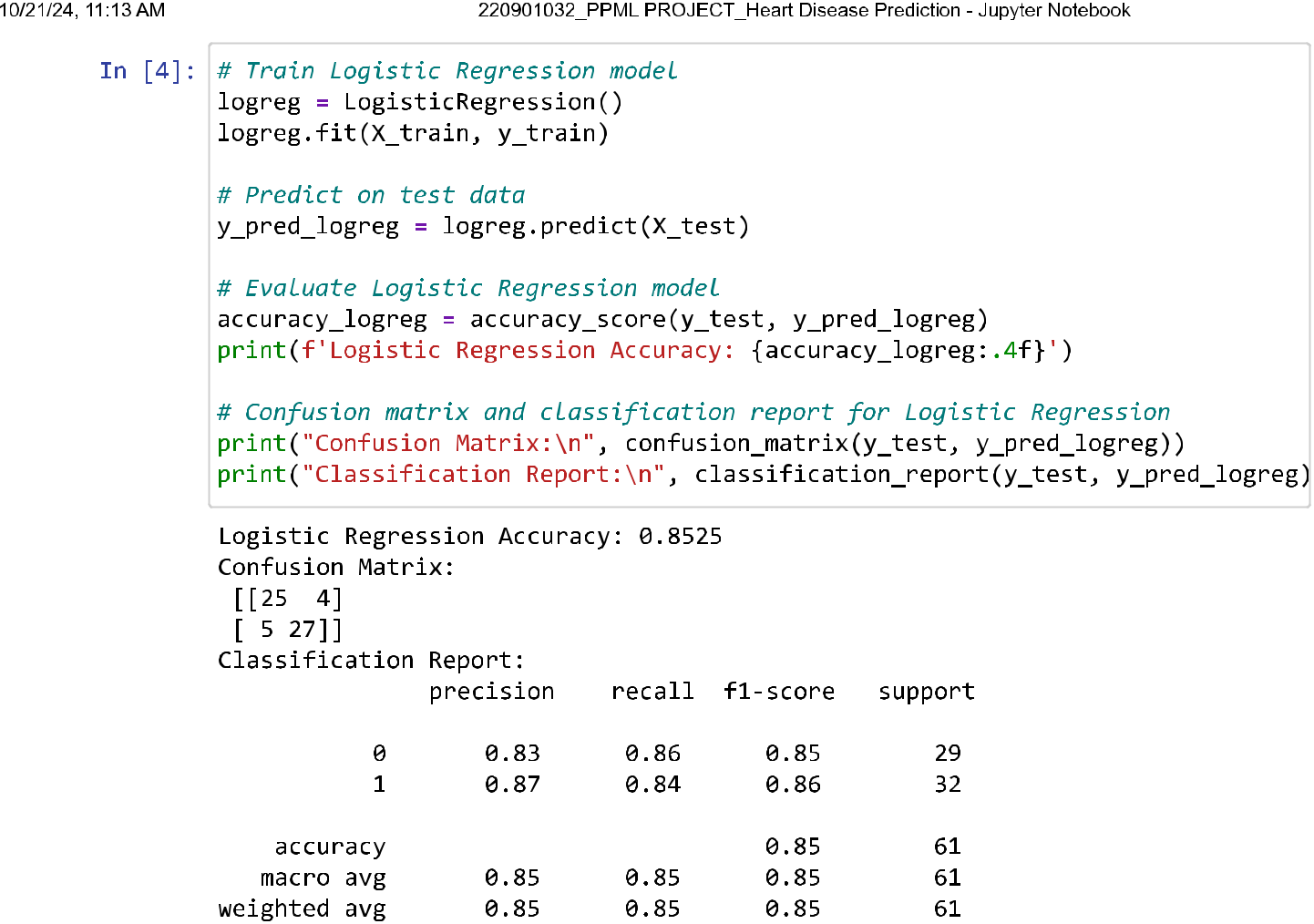
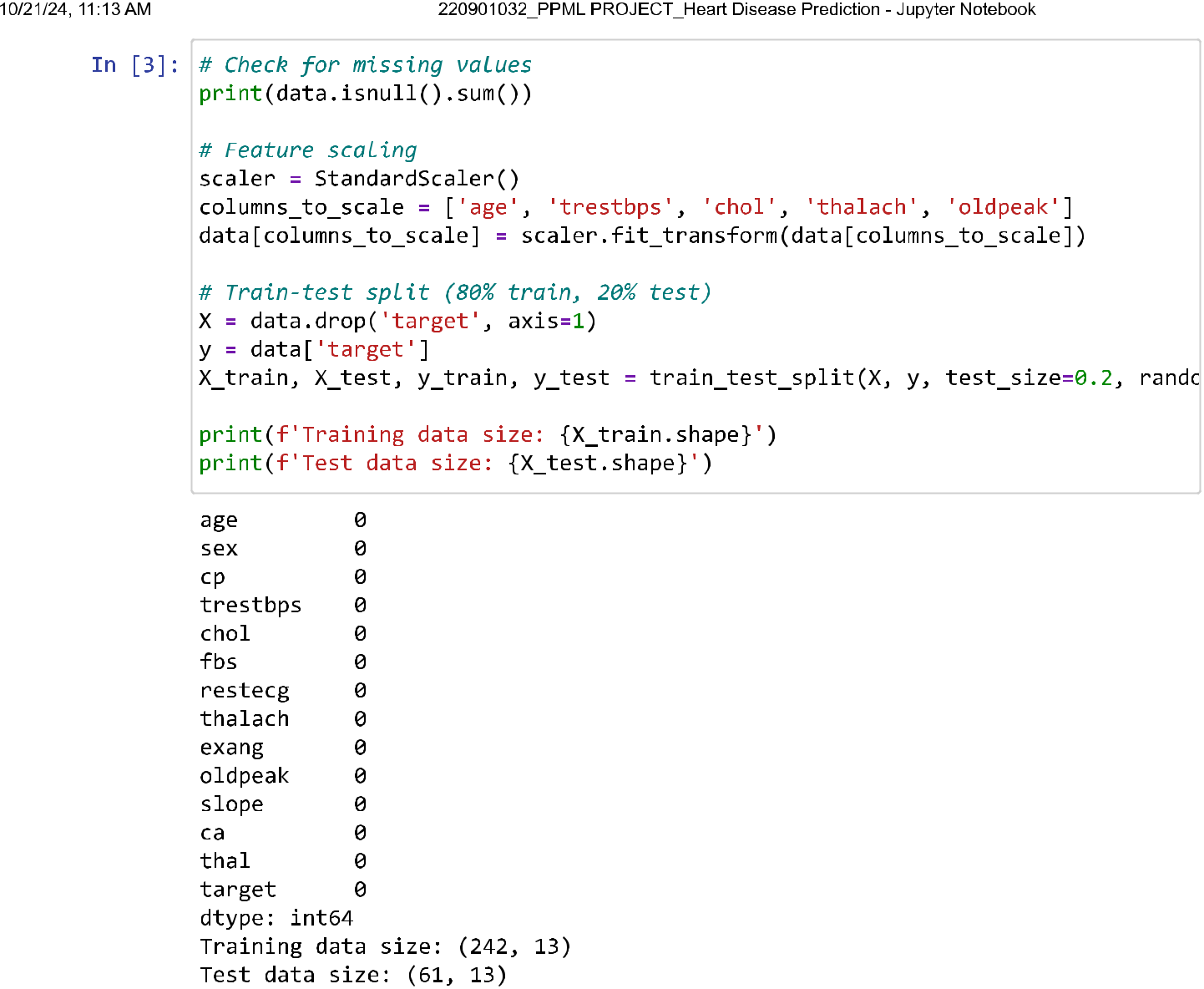
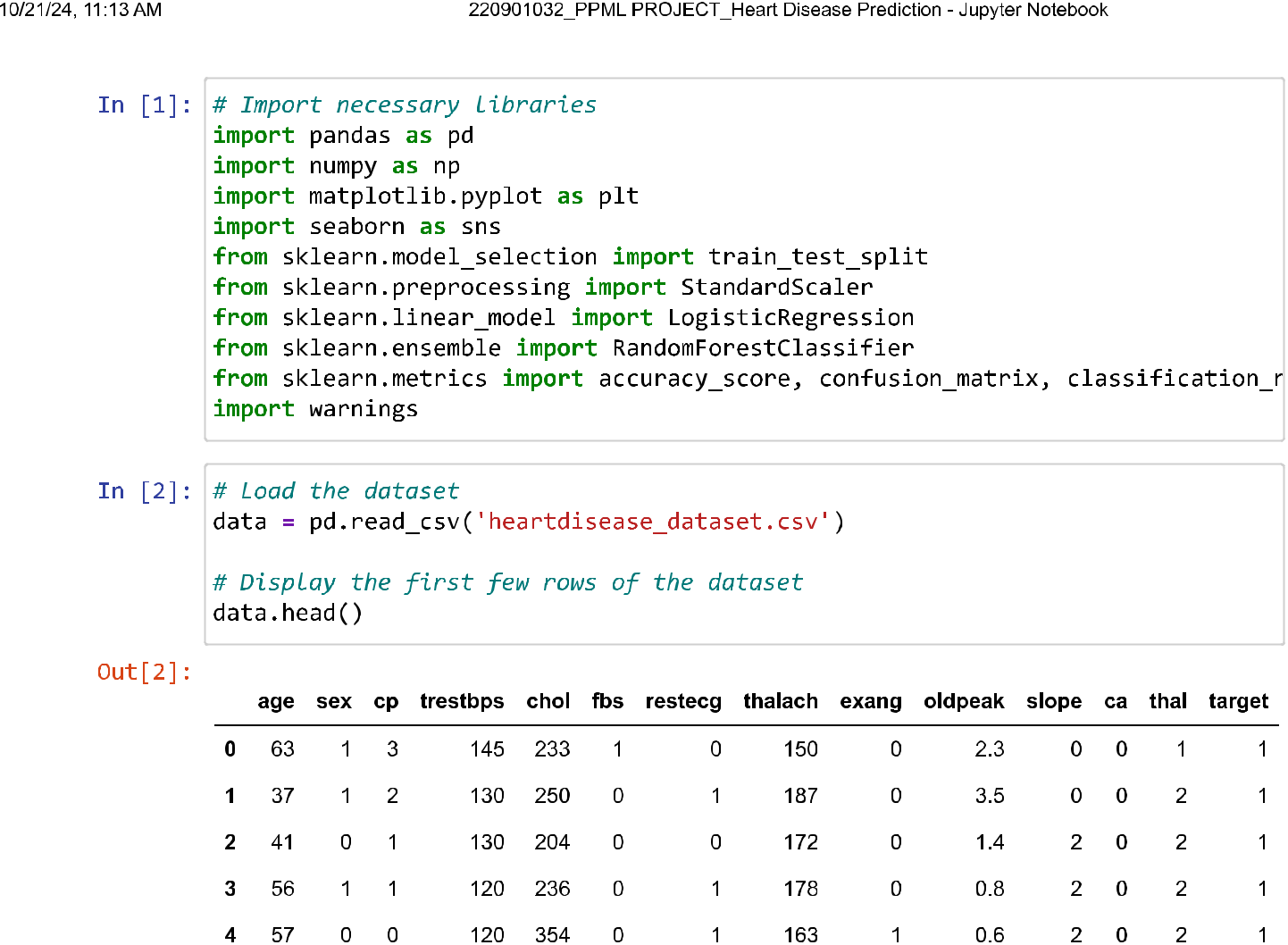
METHODOLOGY

1. **Algorithms Used:**
   * Logistic Regression as a baseline for its simplicity and interpretability.
   * Random Forest for its capability to handle complex, non-linear patterns.
2. **Steps in Model Training:**
   * Data splitting into training and test sets.
   * Training models and hyperparameter tuning.
   * Evaluating models using performance metrics.

IMPLEMENTATION

The code and the CSV File for this project is available on my GitHub repository at:

https://github.com/220901032janarthananm/220901032\_PPML-PROJECT\_Heart-Disease-Prediction



RESULTS AND EVALUATION

**Metrics Used:**

* Accuracy, Confusion Matrix, Precision, Recall, F1-score.

**Key Findings:**

* Logistic Regression achieved moderate accuracy.
* Random Forest provided higher accuracy and meaningful feature importance insights.

**Visualizations:**

* Confusion matrix heatmap, Feature importance bar graph.

FUTURE WORK

**Enhancements:**

* Incorporate larger, more diverse datasets.
* Experiment with deep learning models for improved accuracy.

**Deployment:**

* Develop a user-friendly application for real-time predictions.

FUTURE WORK

The project successfully built a machine-learning model for heart disease prediction, with Random Forest offering superior accuracy.

Insights from feature importance analysis can aid healthcare professionals in prioritizing diagnostic efforts.

This approach demonstrates potential integration into real-world healthcare systems!!!