TITLE: BREAST CANCER PREDICTION USING MACHINE LEARNING

NAME: KELLIE NJOKI NDARU

DATE: 30/08/2024

INTRODUCTION

- **Project Objective**: To develop a model to predict whether a breast tumor is benign or malignant.
- **Dataset Used**: Breast Cancer Wisconsin (Diagnostic) Dataset
- **Stakeholder**: Healthcare providers, patients

PROBLEM STATEMENT

- **Problem**: Early and accurate diagnosis of breast cancer is crucial for effective treatment and patient survival.

- **Goal**: To build a reliable model that can assist in early detection of malignant tumors.

DATA OVERVIEW

- **Dataset description**:
  - 569 samples, 31 features

- **Target Variable**:

Diagnosis (B=Benign, M=Malignant)

DATA PREPROCESSING

Steps taken:

- Dropped irrelevant columns (ID column)
- Handled missing columns (dropped unamed32 column)
- Feature scaling
- Split the data into training and testing sets

MODEL SELECTION

- **Models Evaluated**: Logistic Regression, Decision Tree, Random Forest, Gradient Boosting.

- **Evaluation Metrics**: Accuracy, Precision, Recall, F1-Score.

MODEL PERFORMANCE

- **Logistic Regression**:

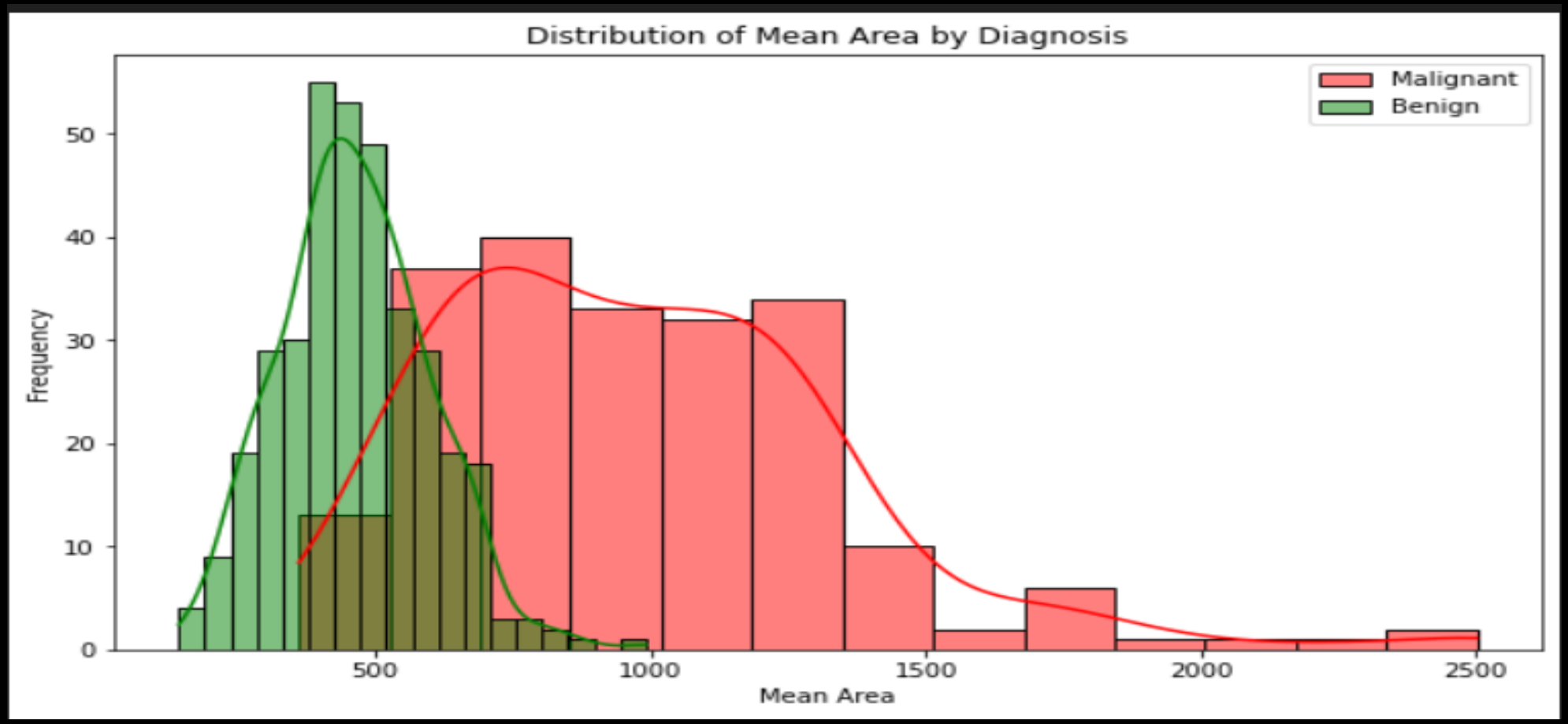
Accuracy: 98%

Precision and Recall: High performance in both classes.

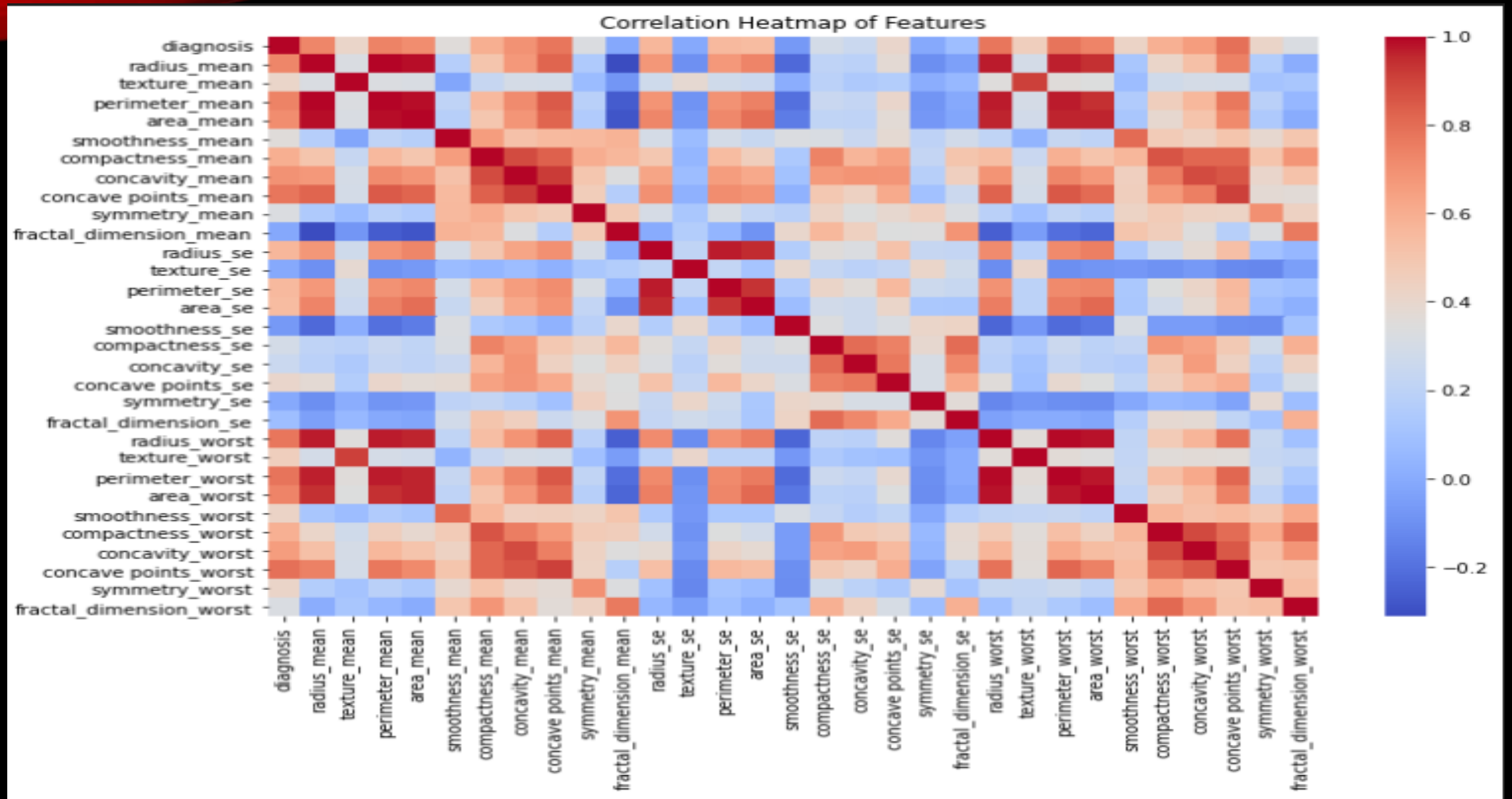
F1-Scores: This indicates a better balance between precision and recall.

VISUALIZATIONS

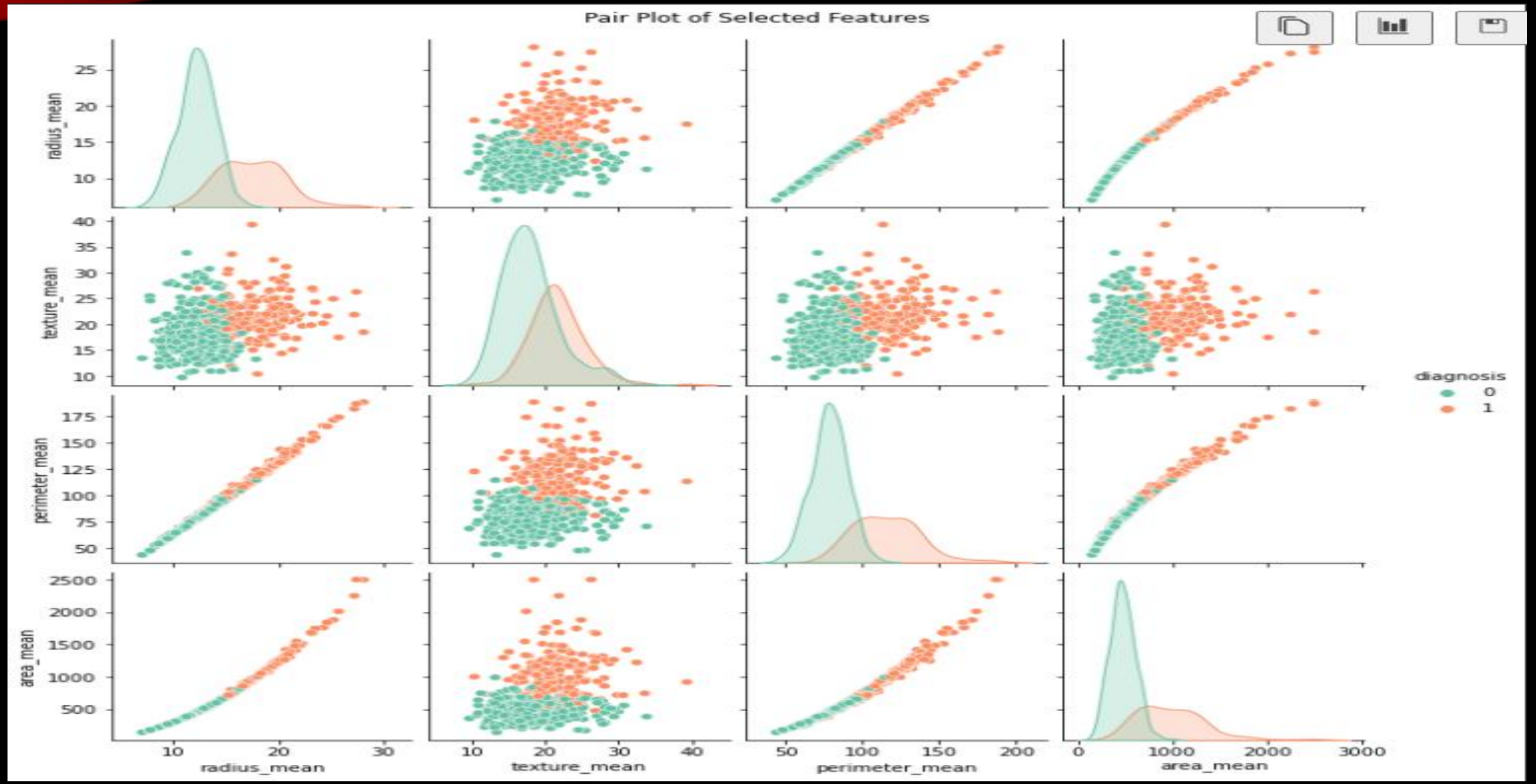
DISTRIBUTION OF FEATURE VALUES FOR EACH DIAGNOSIS



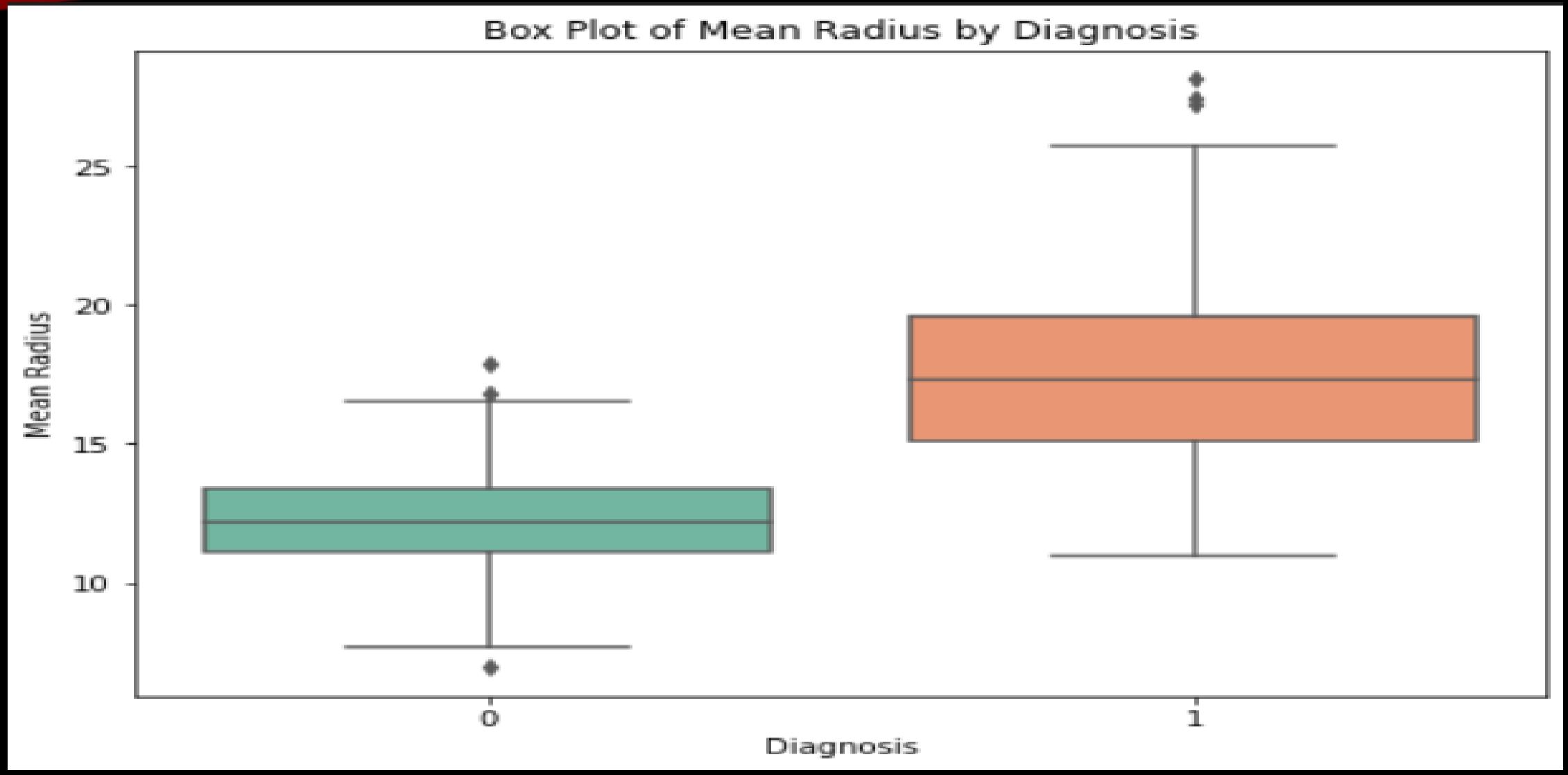
CORRELATION HEATMAP



PAIR PLOT OF SELECTED FEATURES



BOX PLOT OF FEATURE VALUES BY DIAGNOSIS



RECOMMENDATIONS

- Deploy the Logistic Regression Model for real world use.

THANKYOU