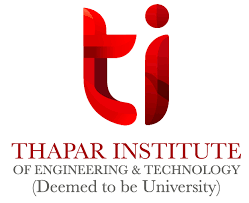
**Heart Issues Detector**



Computer Science and Engineering Department

Thapar Institute of Engineering and Technology

(Deemed to be University), Patiala – 147004

**Machine Learning Project**

Submitted By:

Name of Student: Ravi Kant

Roll No. of Student: 102283013

Submitted To:

Ms. Kudratdeep Aulakh

**Index**

|  |  |  |
| --- | --- | --- |
| **Sr. No.** | **Content used** | **Page No.** |
| 1. | Introduction |  |
| 2 | Libraries used |  |
| 3. | Algorithm(s) used |  |
| 4. | Code and Screenshots |  |

Introduction

Heart disease remains a leading cause of mortality worldwide, making early detection and prevention vital. This project leverages logistic regression , K-nearest neighbour, Random forest classifier , to analyse a dataset containing various health indicators and risk factors. By exploring relationships between these factors and the presence of heart disease, the aim is to build a predictive model. This model can then assess the likelihood of an individual developing heart disease based on their unique combination of risk factors. Ultimately, the goal is to create a reliable tool that assists healthcare professionals in early identification and intervention.

Problem Statement

Developing a predictive model using logistic regression,K-nearest neighbor and Random forest classifier, to accurately identify the presence or absence of heart disease based on various health indicators and risk factors. The objective is to create a tool that aids healthcare professionals in early detection and assessment of an individual's risk of developing heart disease, thereby facilitating timely interventions and improving patient outcomes.

Link for dataset : [Dataset Link](file:///C:\Users\RIJUL%20BHATT\Desktop\ML%20project\heart.csv)

Algorithm Used

The methodology for predicting heart disease using logistic regression typically involves several key steps:

1. Data Collection and Preprocessing:

Gather a comprehensive dataset containing relevant features such as age, gender, cholesterol levels, blood pressure, etc.

Preprocess the data, handling missing values, encoding categorical variables, and normalizing/standardizing numerical features.

1. Exploratory Data Analysis (EDA):

Analyse the dataset to understand relationships between different variables and their distributions.

Identify correlations and patterns that might indicate potential risk factors for heart disease.

1. Feature Selection:

Use statistical techniques or domain knowledge to select the most relevant features that strongly correlate with the presence of heart disease.

Remove irrelevant or redundant features to enhance model performance.

1. Splitting Data into Training and Testing Sets:

Divide the dataset into training and testing subsets. The training set is used to train the model, while the testing set evaluates its performance.

1. Logistic Regression Model Building:

Apply logistic regression, a classification algorithm suitable for binary outcomes (presence or absence of heart disease).

Train the model on the training dataset, adjusting parameters to optimize its performance.

Validate the model on the testing dataset to ensure its generalizability and assess its predictive capabilities.

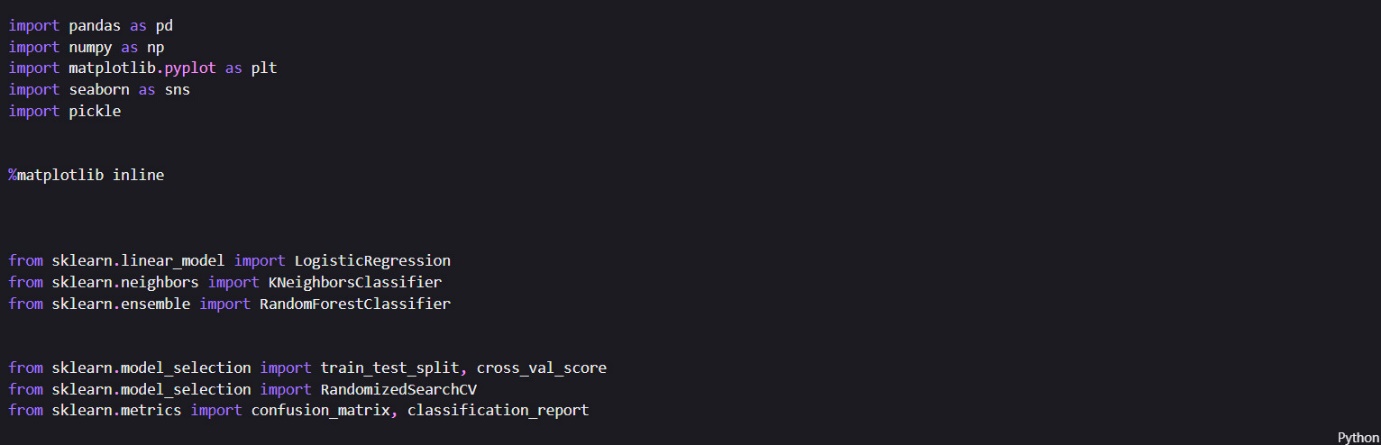
1. K-nearest neighbor Model Building:

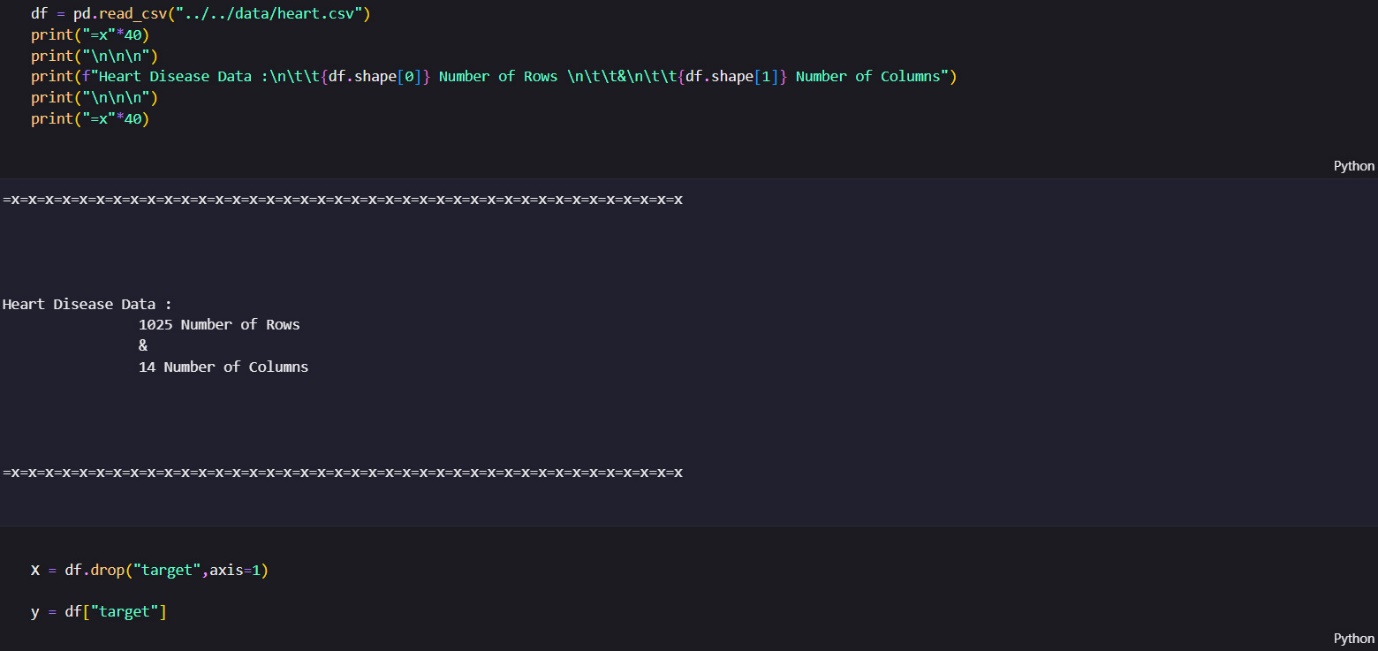
This code is using KNN with different numbers of neighbors to find the optimal number that maximizes the testing score. It iterates through a range of neighbor values, records the training and testing scores for each, and then sets the number of neighbors to the one that gave the highest testing score before finally retraining the model with that optimal number of neighbors. This is a common approach to hyperparameter tuning for KNN models.

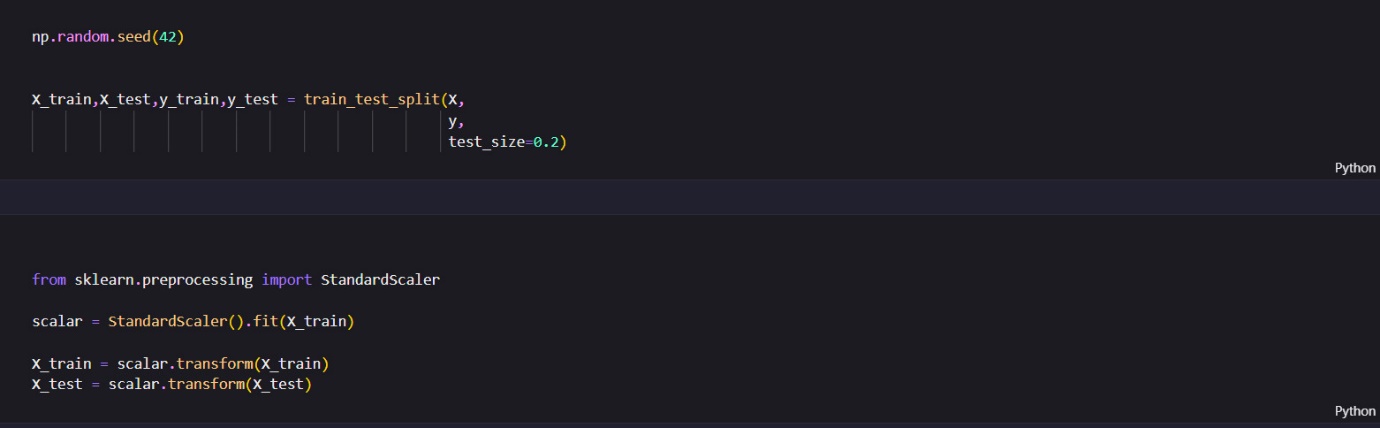
1. Random forest classifier Model Building:

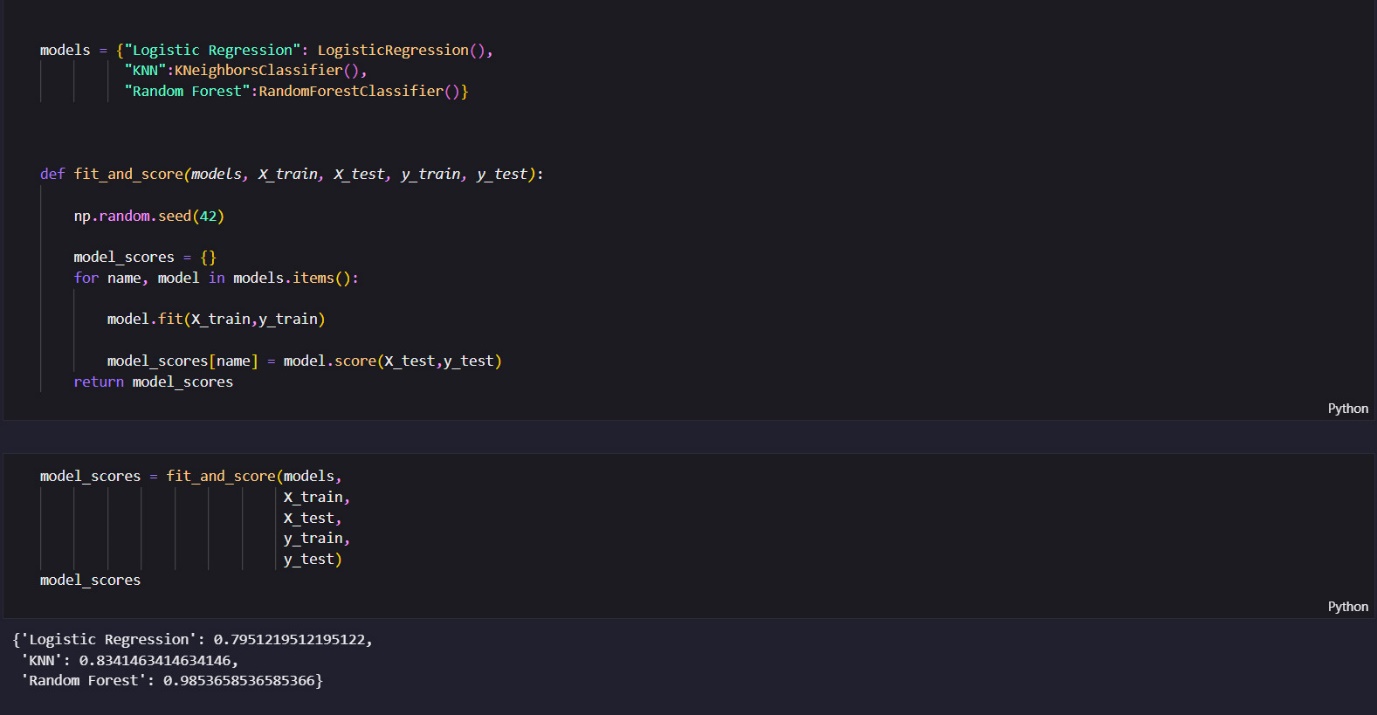
The Random Forest algorithm is employed for its ensemble nature, combining multiple decision trees to make robust predictions. The workflow involves training the model on a subset of the data and then using it to predict outcomes on new, unseen data. The code provides a foundational understanding of how to use Random Forest for classification tasks.

Results

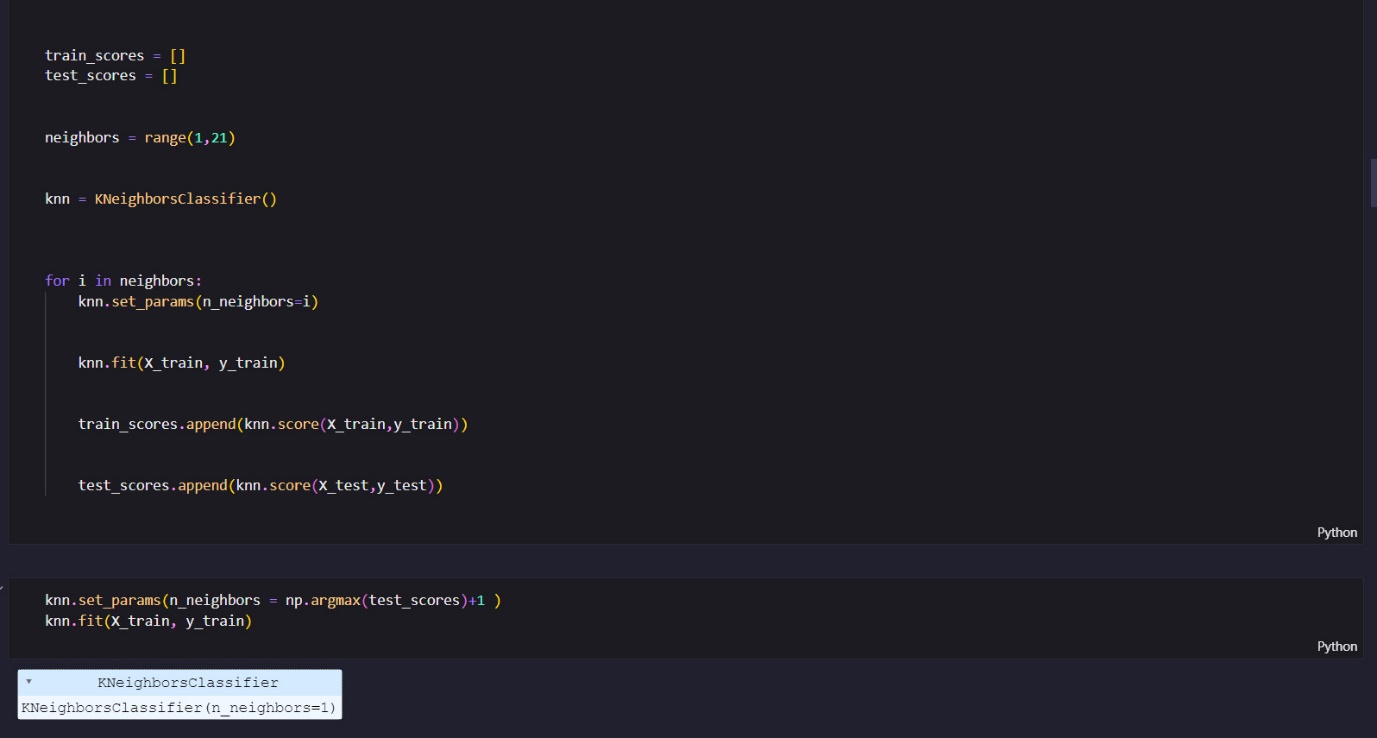


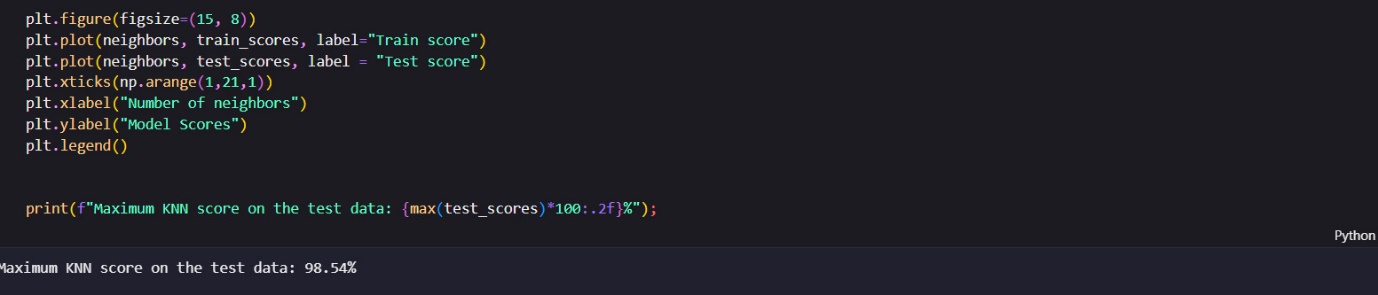


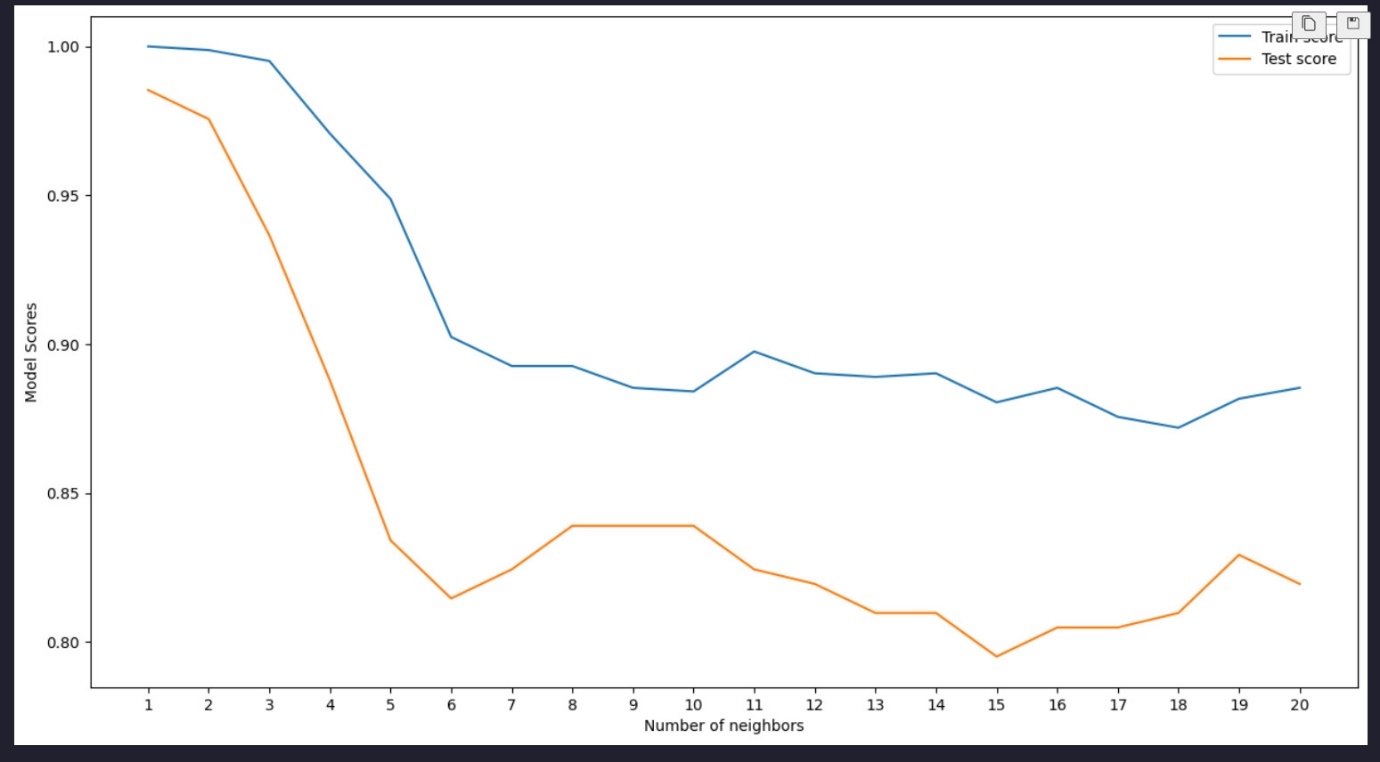


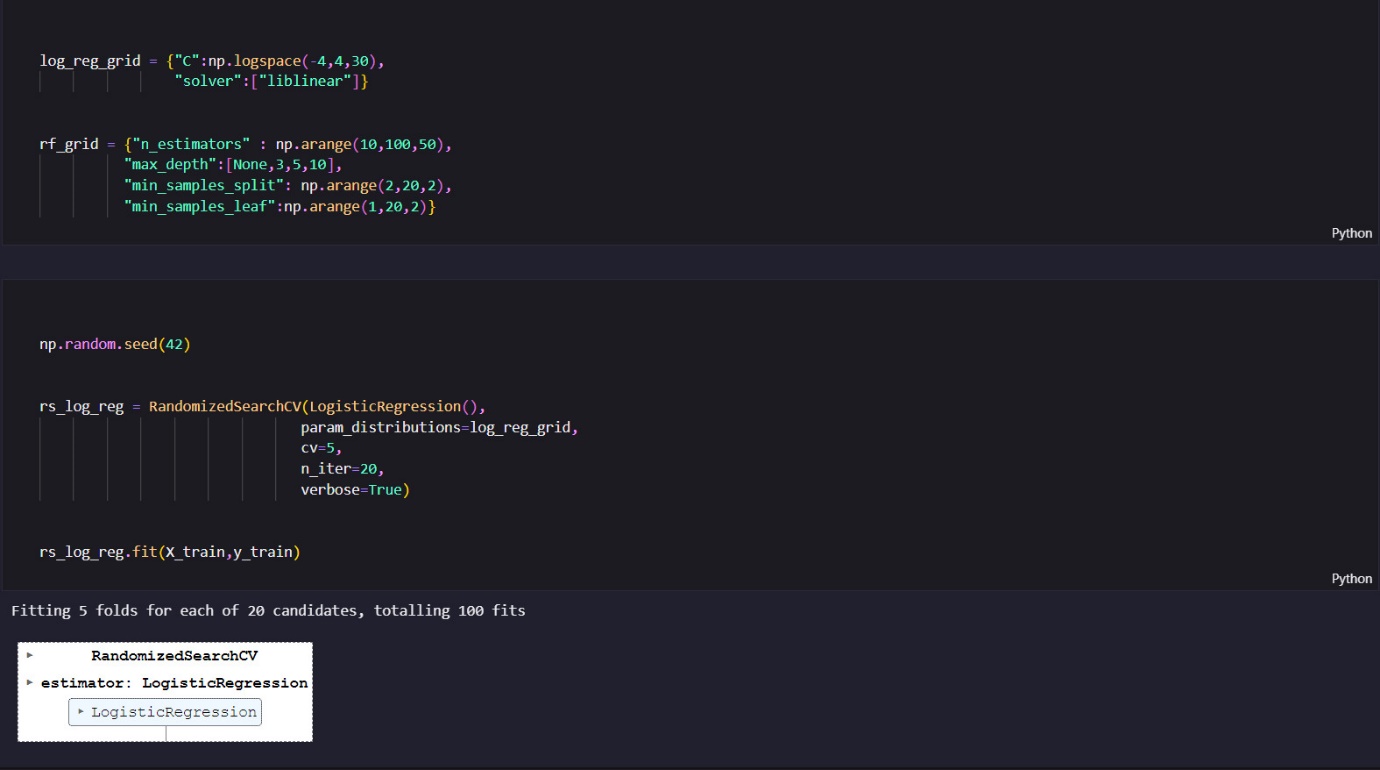


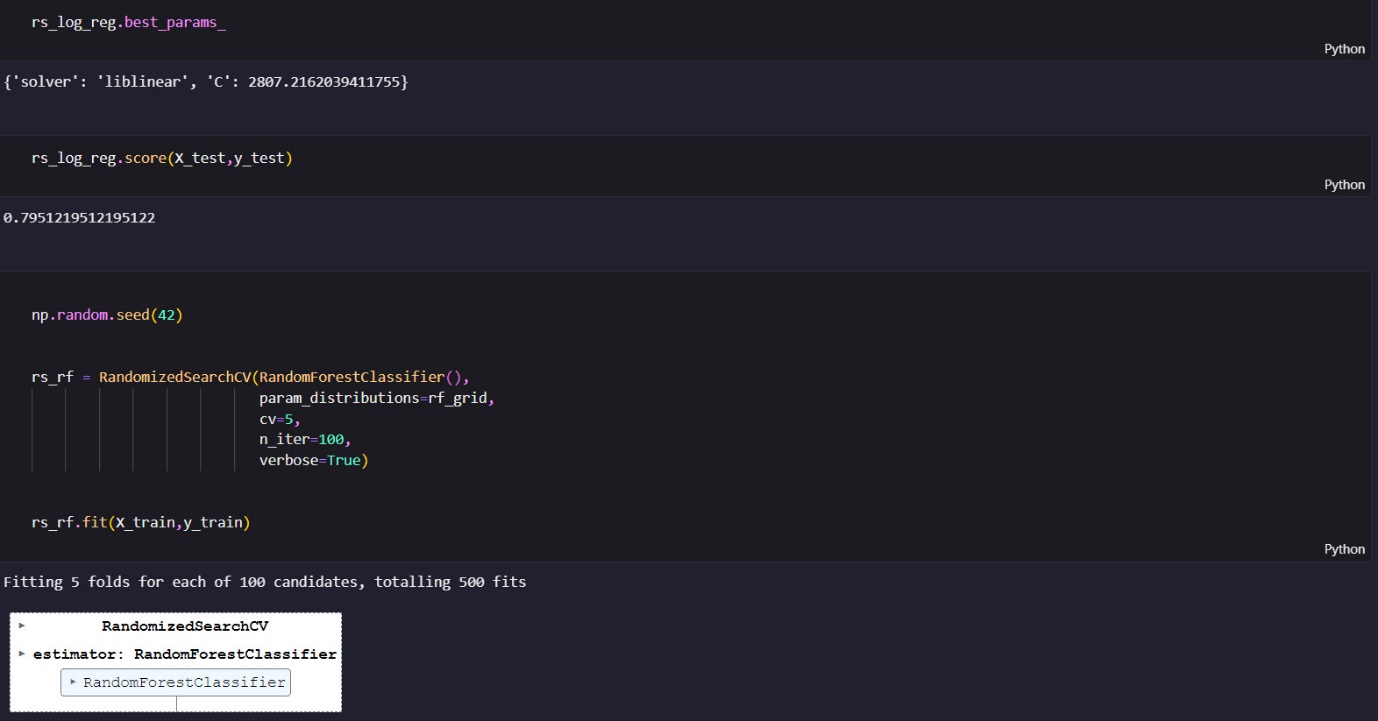


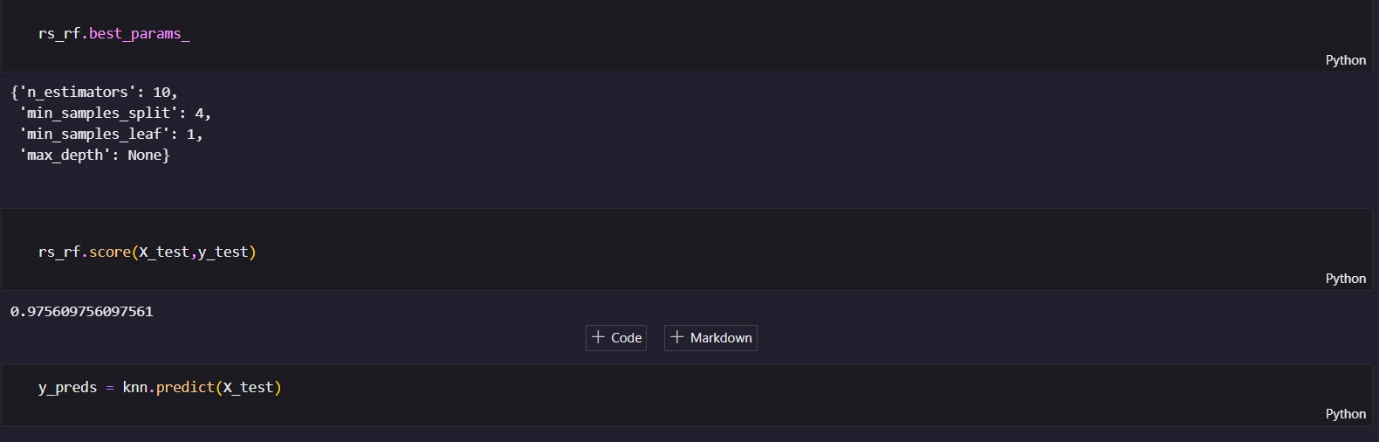


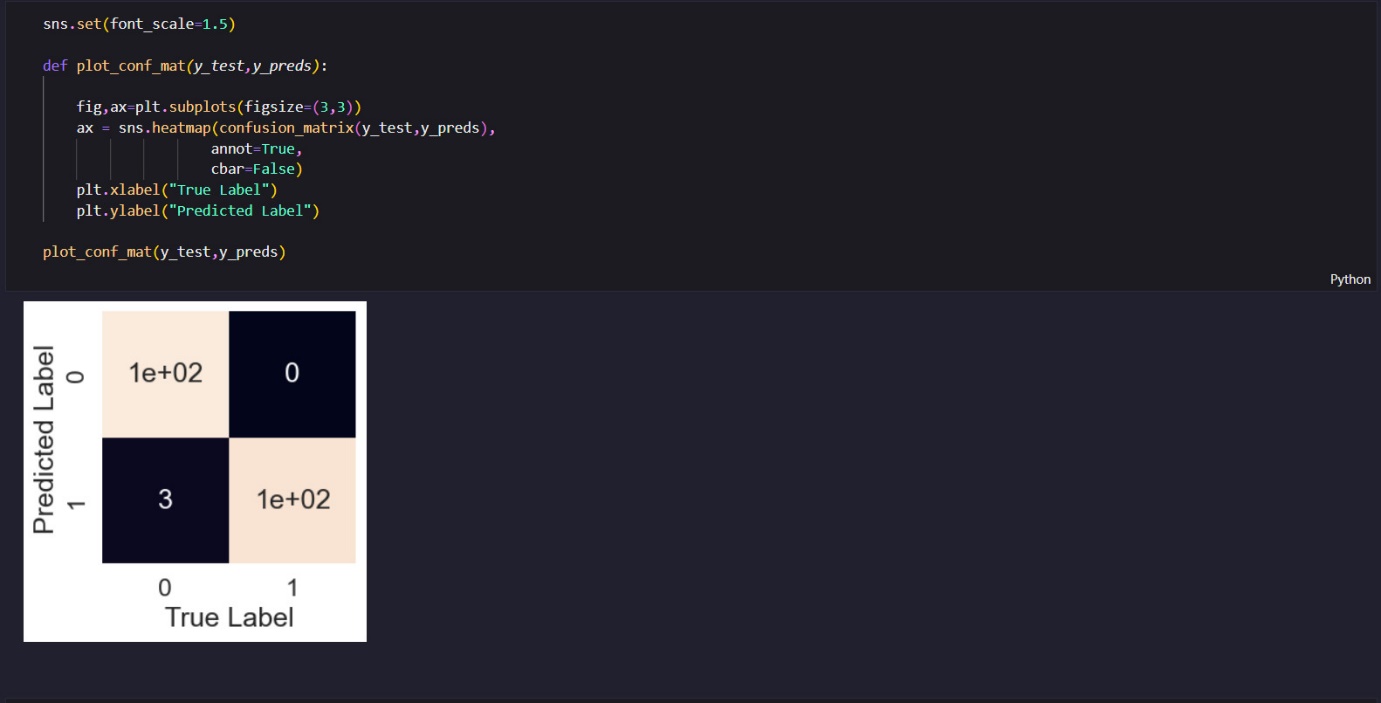


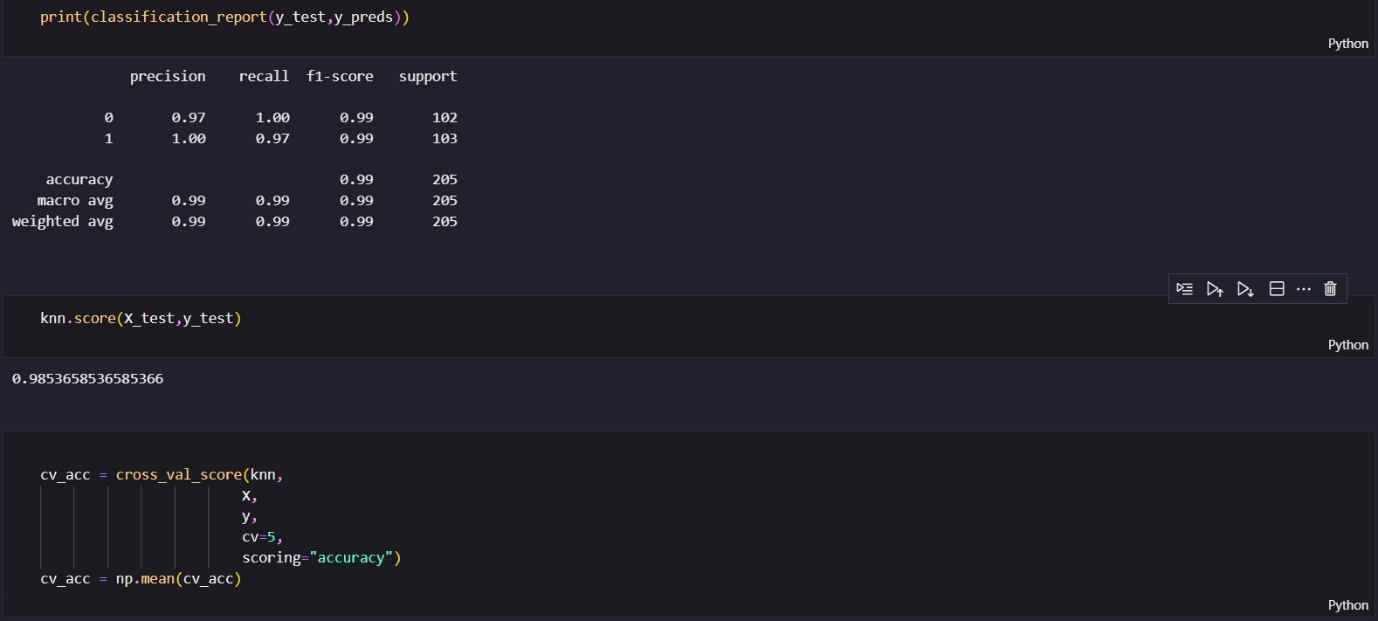


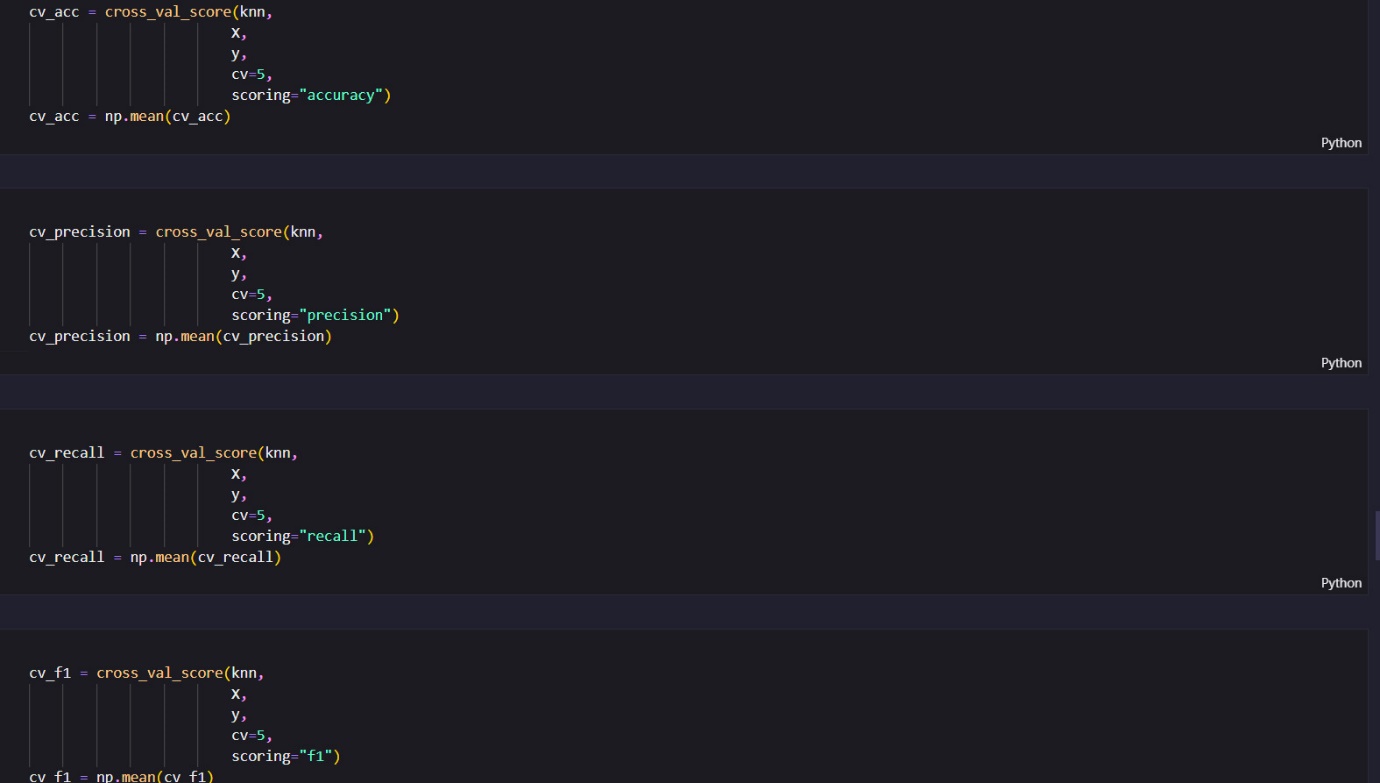




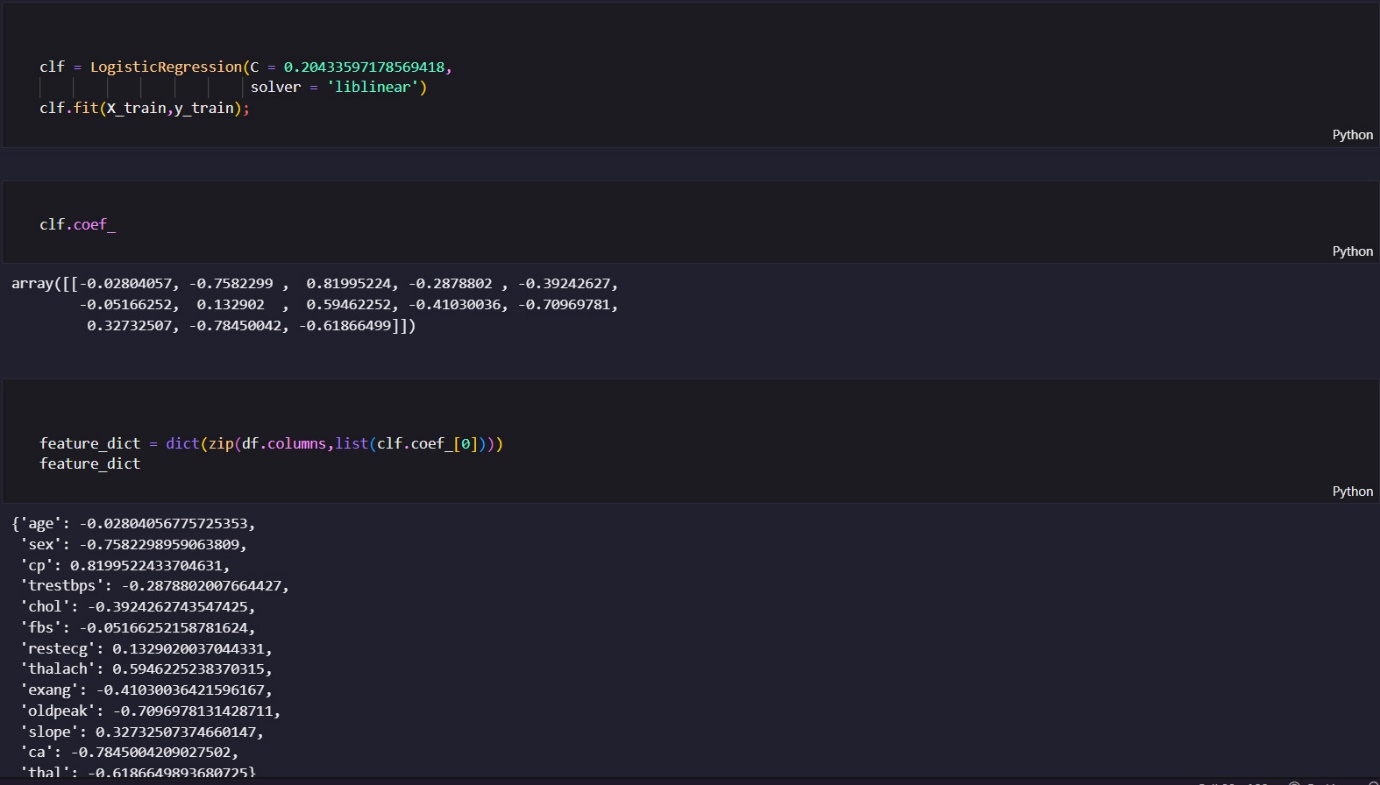




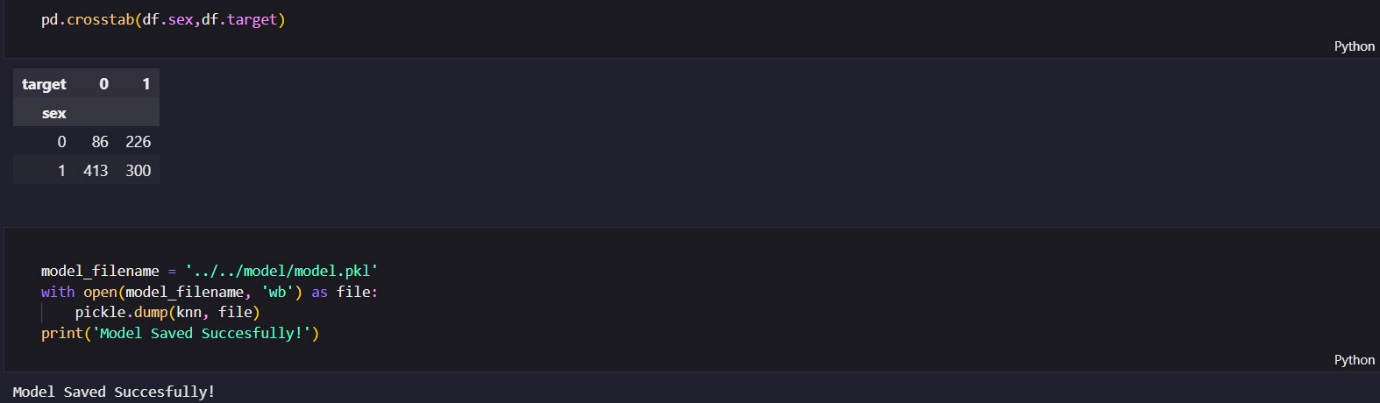












Conclusion

My project built a tool using math and data to predict if someone might have heart problems. We looked at lots of info like age, BMI, heartrate, lifestyle etc, to make this prediction. Our tool did pretty well when we tested it, showing it could predict heart issues accurately.

This tool could be super helpful for doctors to catch heart problems early and help people stay healthier. But, we still need more tests and info to be totally sure it's always right. So, while it's a good start, we can keep working to make it even better and more useful for helping people stay heart-healthy.