# University of Petroleum and Energy Studies (UPES), Dehradun , Dehradun ...

# Elements of AI ML (Assignment-1)

Name -Devansh Saini

Sap id – 500119031

Roll no – R2142230351

Batch – B11 Btech.CSE

## Heart Disease Prediction Model

--Based on Sdg 3

**1. Introduction**

This project is designed to predict the likelihood of heart disease using machine learning models. With heart disease as a leading cause of death globally, the ability to predict risk factors early could lead to timely intervention, potentially saving lives.

**2. Purpose**

The main purpose of this model is to assist healthcare providers in identifying patients at higher risk for heart disease, enabling them to take preventative actions early and optimize resource allocation.

**3. Motivation for Choosing the Topic**

With heart disease affecting millions worldwide, finding ways to leverage data science for early diagnosis is a critical area of interest. This project explores how predictive modelling could contribute to proactive healthcare measures, making it a valuable endeavour.

**4. Dataset Used**

The Cleveland Heart Disease dataset, widely used for heart disease prediction, was chosen for this project. It includes 297 patient records with features like age, cholesterol levels, blood pressure, and several other key indicators relevant to heart disease risk assessment.

**5. Libraries Used**

We utilized the following libraries in our project:

* **Pandas** for data manipulation and analysis.
* **NumPy** for numerical operations.
* **Scikit-learn** for model implementation, evaluation, and data preprocessing.
* **Joblib** for model persistence (saving the best model).
* **Google Collab Drive** for easy file handling and saving results directly to Google Drive.

**6. Methodology**

Our approach involved several key steps:

**a. Data Preprocessing**

* **Handling Missing Values**: The dataset had no missing values, allowing us to proceed without imputing data.
* **Feature Encoding**: Converted categorical variables to numeric form for model compatibility, for example, converting 'sex' to binary (1 for male, 0 for female).
* **Feature Scaling**: Standardized features to ensure uniformity, improving model performance and convergence.

**b. Feature Selection**

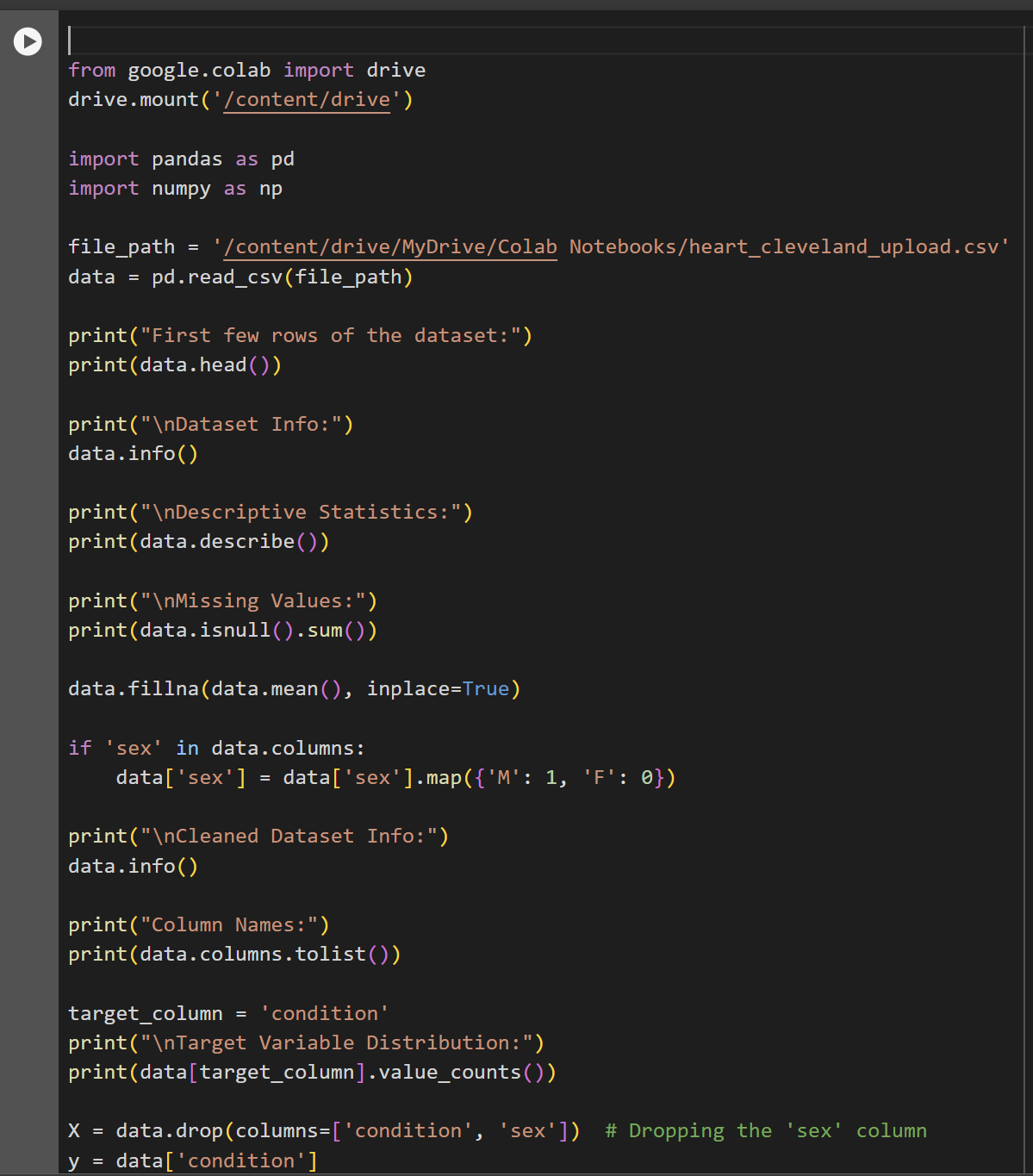
* By analysing correlations between features and the target variable, we retained the most relevant features for heart disease prediction.

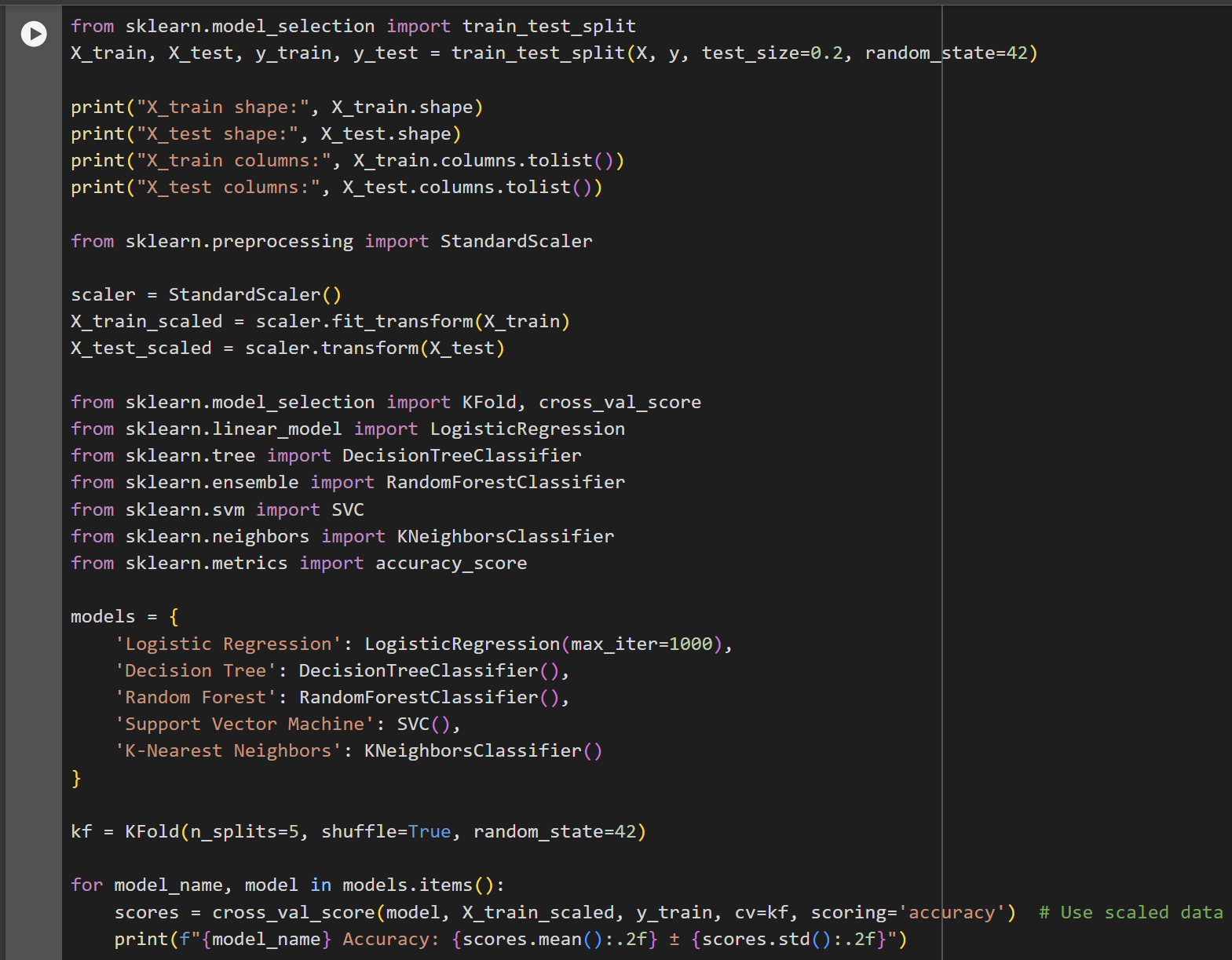
**c. Model Selection and Evaluation**

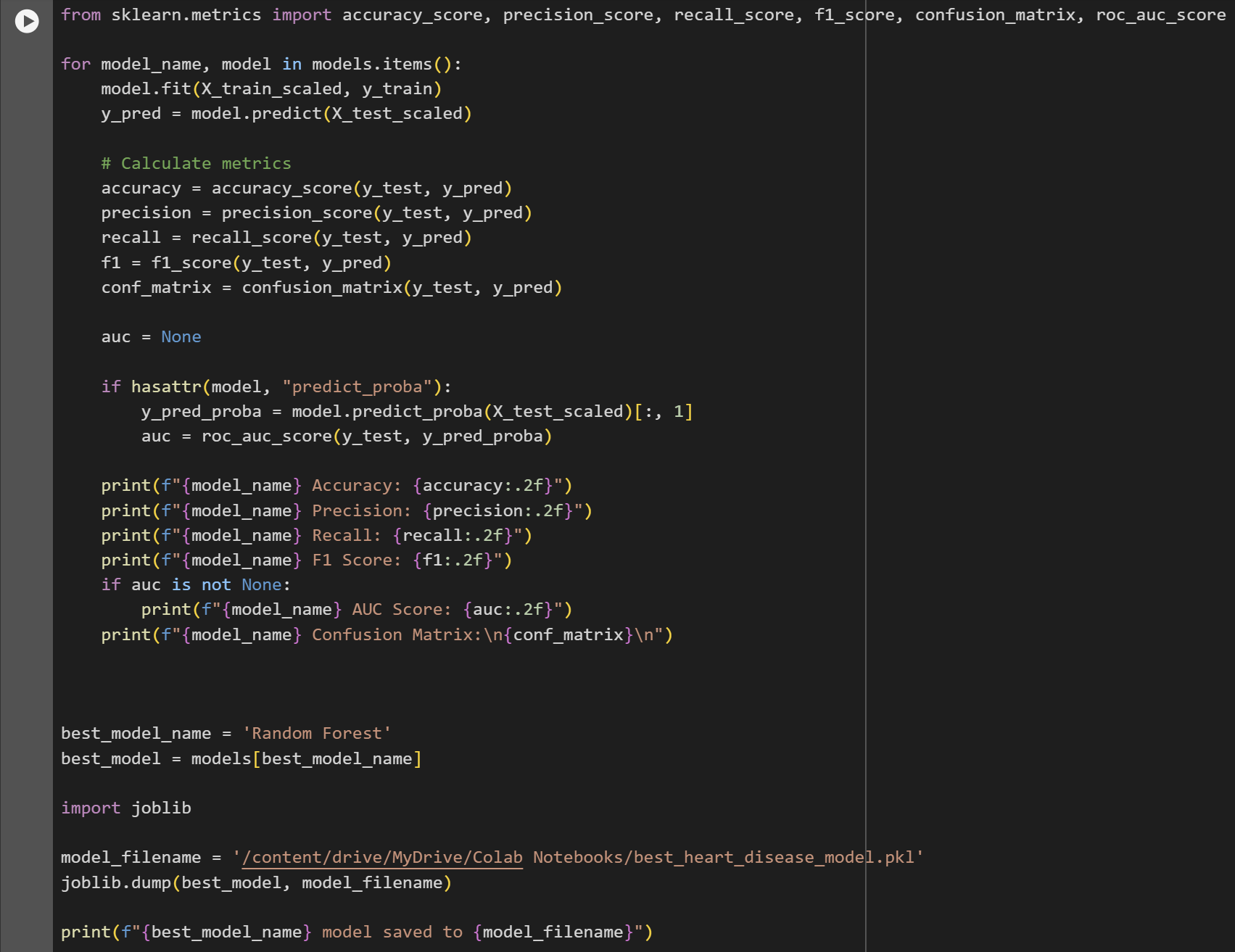
* We implemented five machine learning models: **Logistic Regression**, **Decision Tree**, **Random Forest**, **Support Vector Machine (SVM)**, and **K-Nearest Neighbours (KNN)**.
* **K-Fold Cross-Validation**: We applied 5-fold cross-validation for model evaluation, providing a more robust assessment of model performance.
* **Metrics Evaluated**: Accuracy, precision, recall, F1 score, and AUC were evaluated to determine the best-performing model.

**d. Model Persistence**

* **Saving the Best Model**: Based on the evaluation metrics, Random Forest performed the best. We saved this model using Joblib for easy reuse in future predictions.

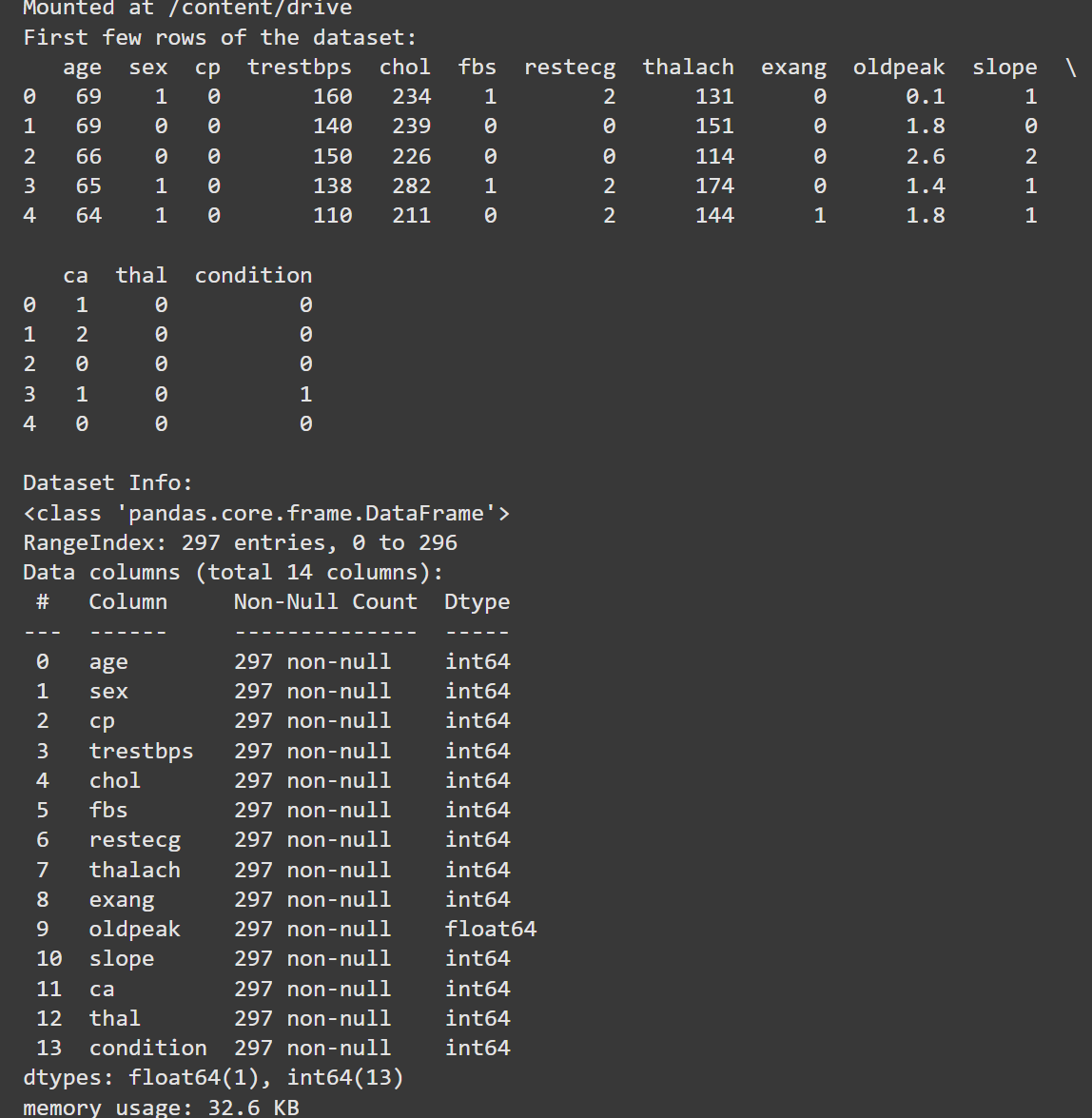


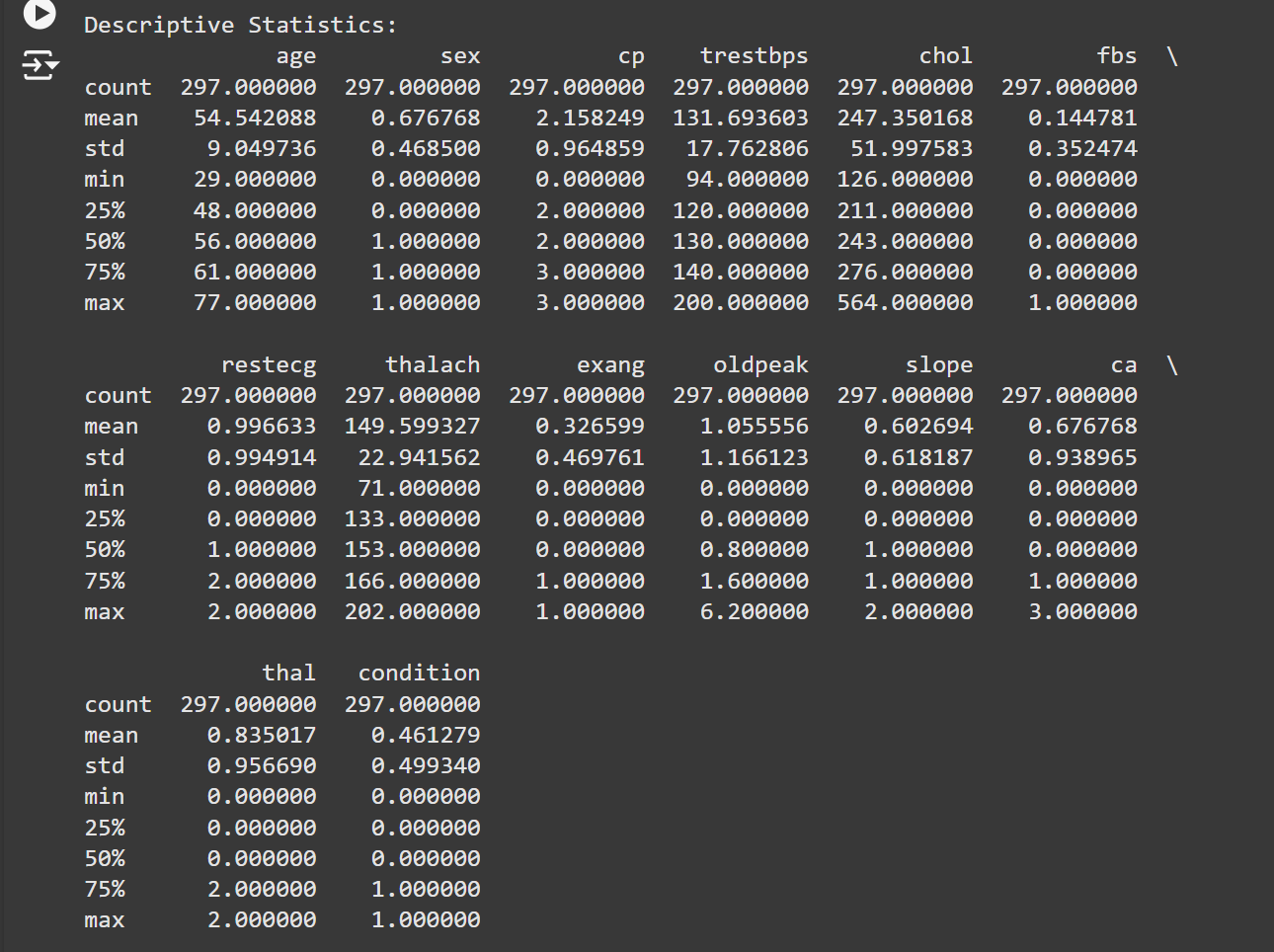


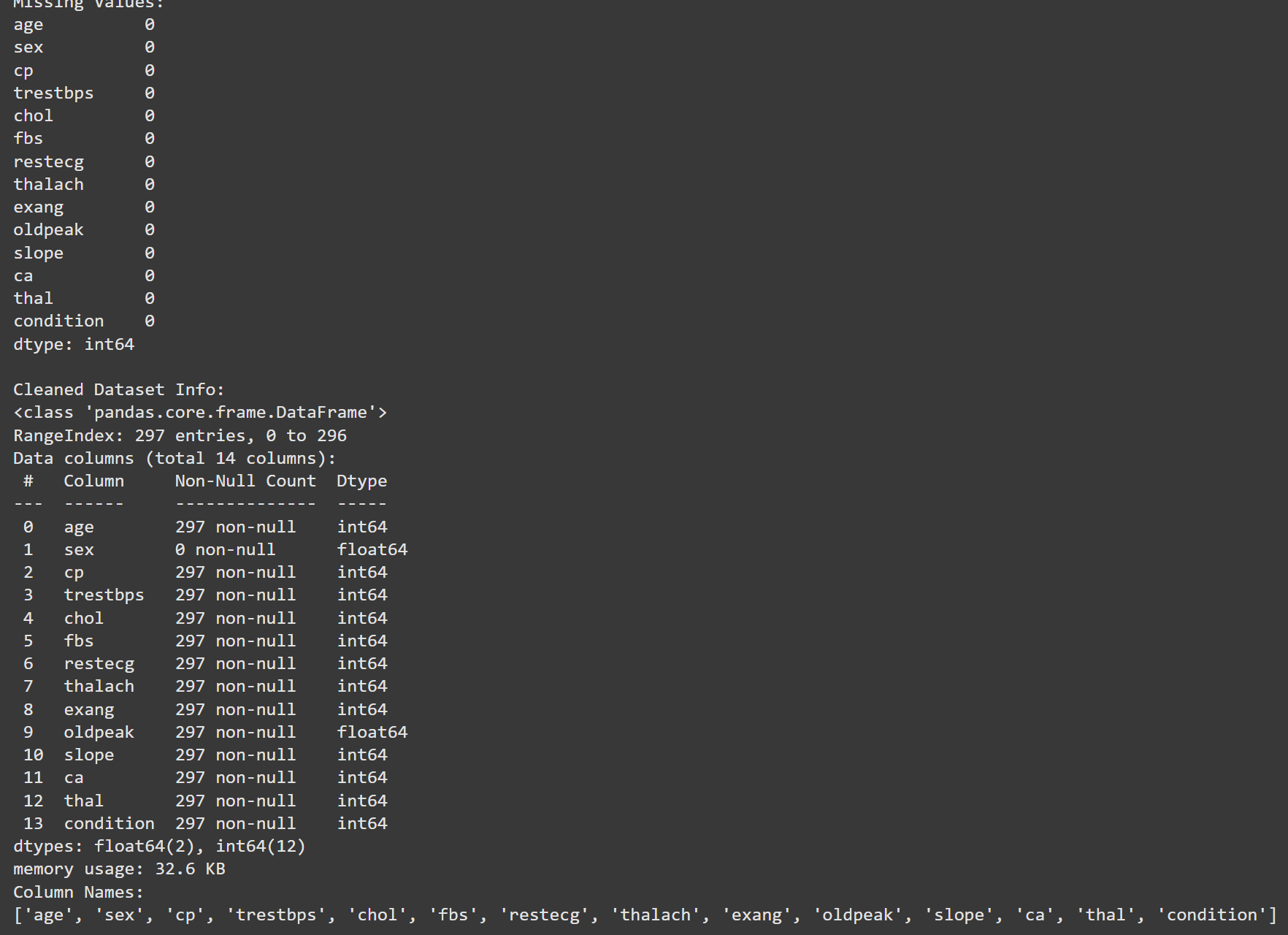


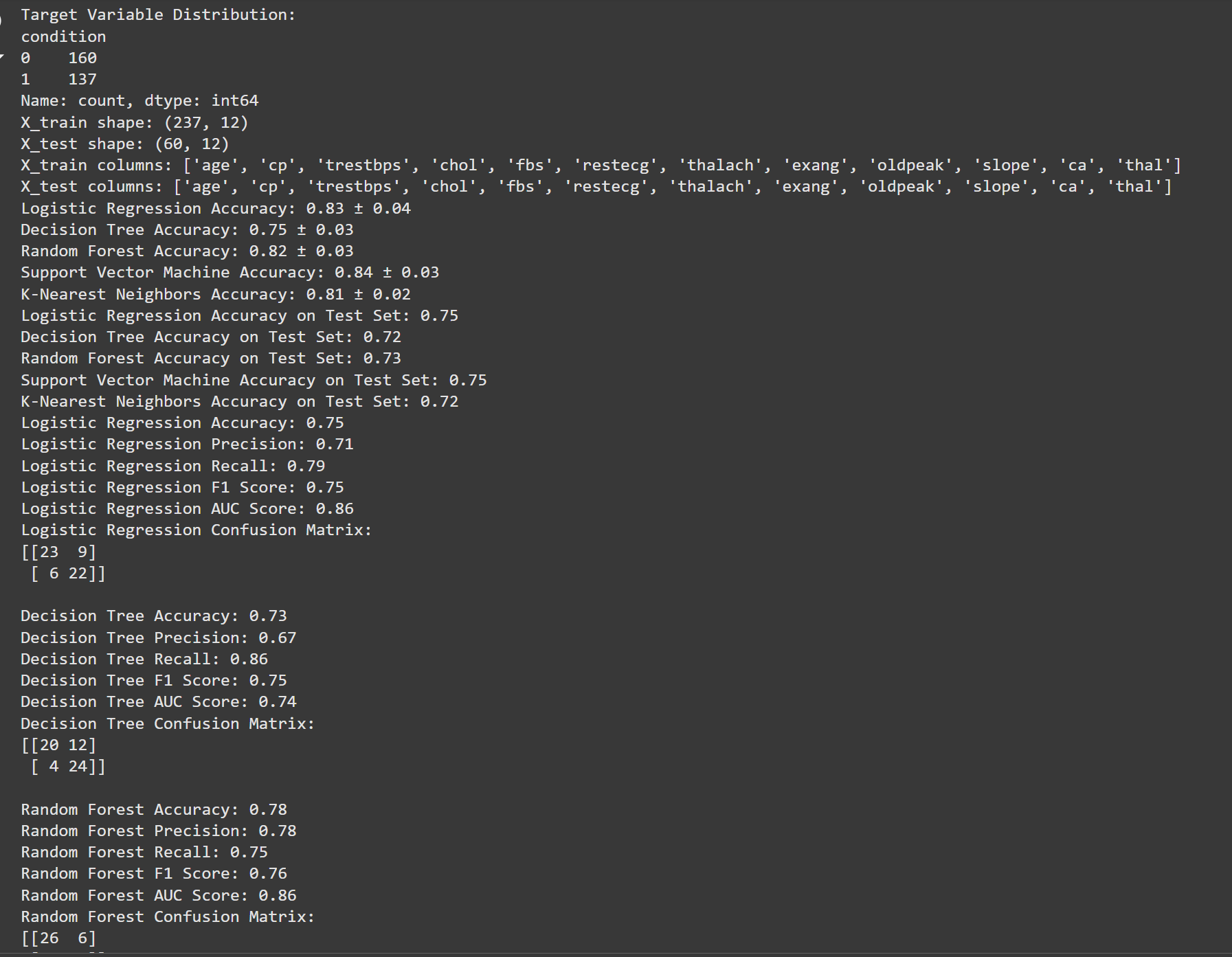
**7. Results**

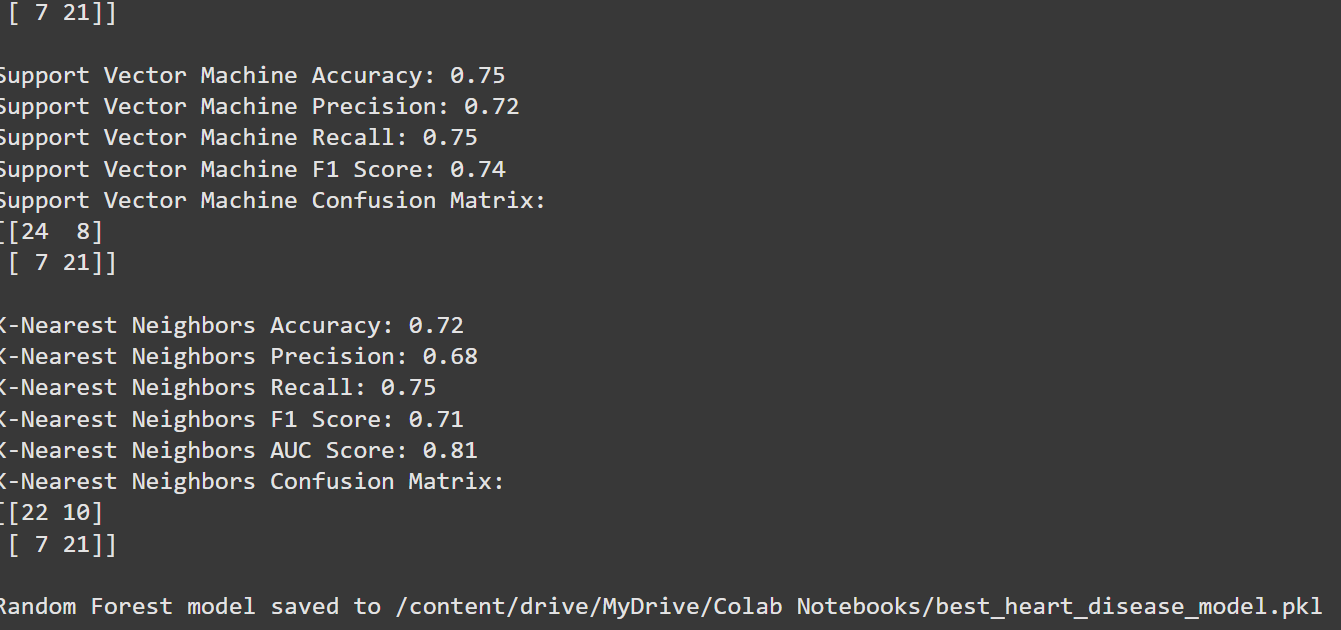
After evaluating the models, Random Forest achieved the highest accuracy (78%) on the test set, with favourable precision, recall, and AUC scores. This model demonstrated the best combination of metrics, suggesting its efficacy for heart disease risk prediction.











**8. Summary**

This project successfully demonstrates the application of machine learning in healthcare, specifically for heart disease risk prediction. While the results are promising, future work could explore larger datasets and additional features to improve accuracy further and expand the model’s real-world applicability.