Heart Disease Prediction Using Machine Learning

1. Objective:

To develop a machine learning model that can predict whether a patient has heart disease based on medical attributes using the Cleveland Heart Disease dataset.

---

2. Dataset Used:

- **File:** `heart_cleveland_upload.csv` (From Kaggle)
- **Source:** Contains 303 samples and 14 medical attributes.
- **Target Column:** `target`
    - `0`: No Heart Disease
    - `1`: Presence of Heart Disease

---

3. Key Features:

- Age, Sex, Chest Pain Type (`cp`), Resting Blood Pressure (`trestbps`), Cholesterol, Fasting Blood Sugar, Resting ECG (`restecg`), Max Heart Rate (`thalach`), Exercise Induced Angina (`exang`), ST Depression (`oldpeak`), Slope, Number of Major Vessels (`ca`), and Thalassemia (`thal`).

---

4. Preprocessing Steps:

- Handled missing values.
- Performed **one-hot encoding** on categorical features (`cp`, `restecg`, `slope`, `thal`).
- Applied **MinMaxScaler** on numeric columns for consistent model input.
- Split data into training and testing sets (80/20).

---

5. Models Trained:

- **Logistic Regression**
- **K-Nearest Neighbors (KNN)**
- **Support Vector Machine (SVM)**
- **Decision Tree**
- **Random Forest**

---

6. Model Evaluation Metric:

- Used **F1 Score** to evaluate model performance.

---

7. Best Model:

- **Random Forest Classifier**
  - Achieved the **highest F1 Score** on the test data.
  - Chosen as the final model for making predictions.

---

8. Final Prediction Test:

- A sample/fake patient's data was input.
- Preprocessed in the same way as training data.
- Predicted using the trained Random Forest model.

Example Output:

```
Predicted Class (0 = No Disease, 1 = Disease): 1
Prediction Probability: [0.42 0.58]
```

---

9. Conclusion:

The Random Forest model provided the most accurate results in predicting heart disease based on patient attributes. This tool can help in early diagnosis and improve patient care decisions.
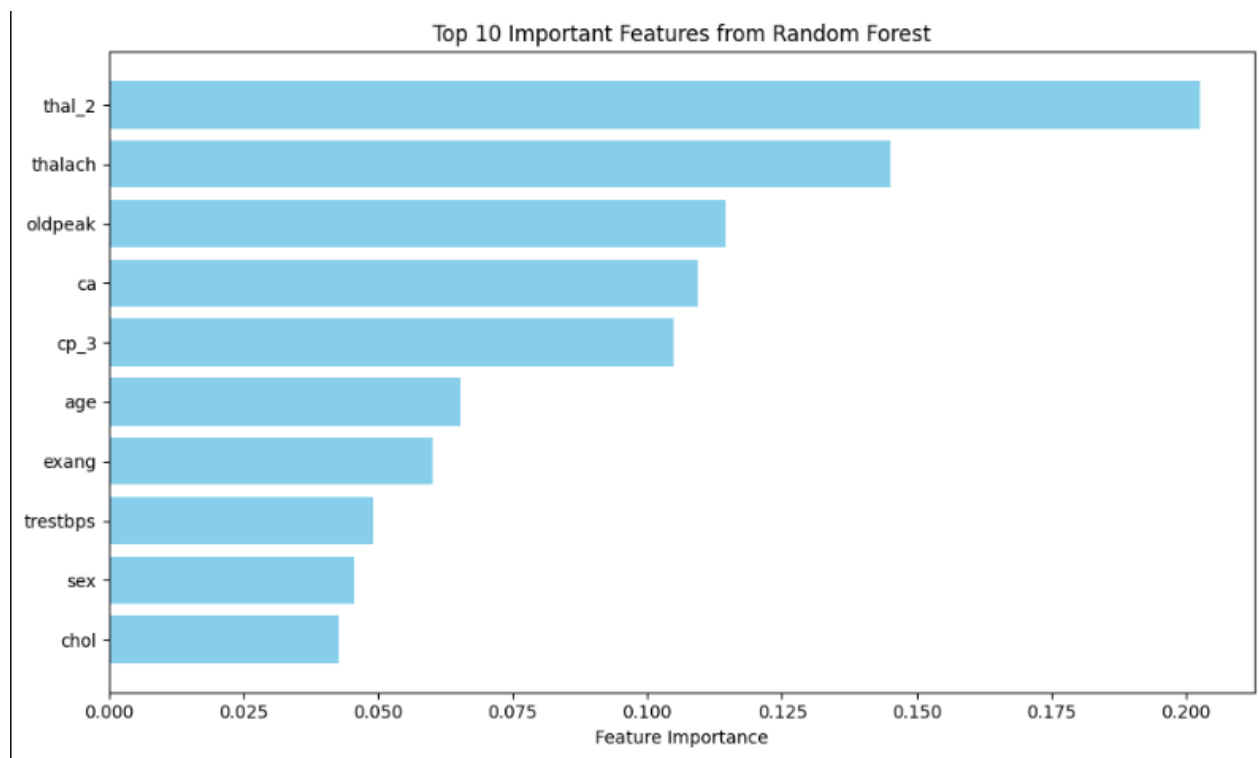
Visual Explorations & Results

1. Heatmap showing correlation between variables.

Correlation Heatmap
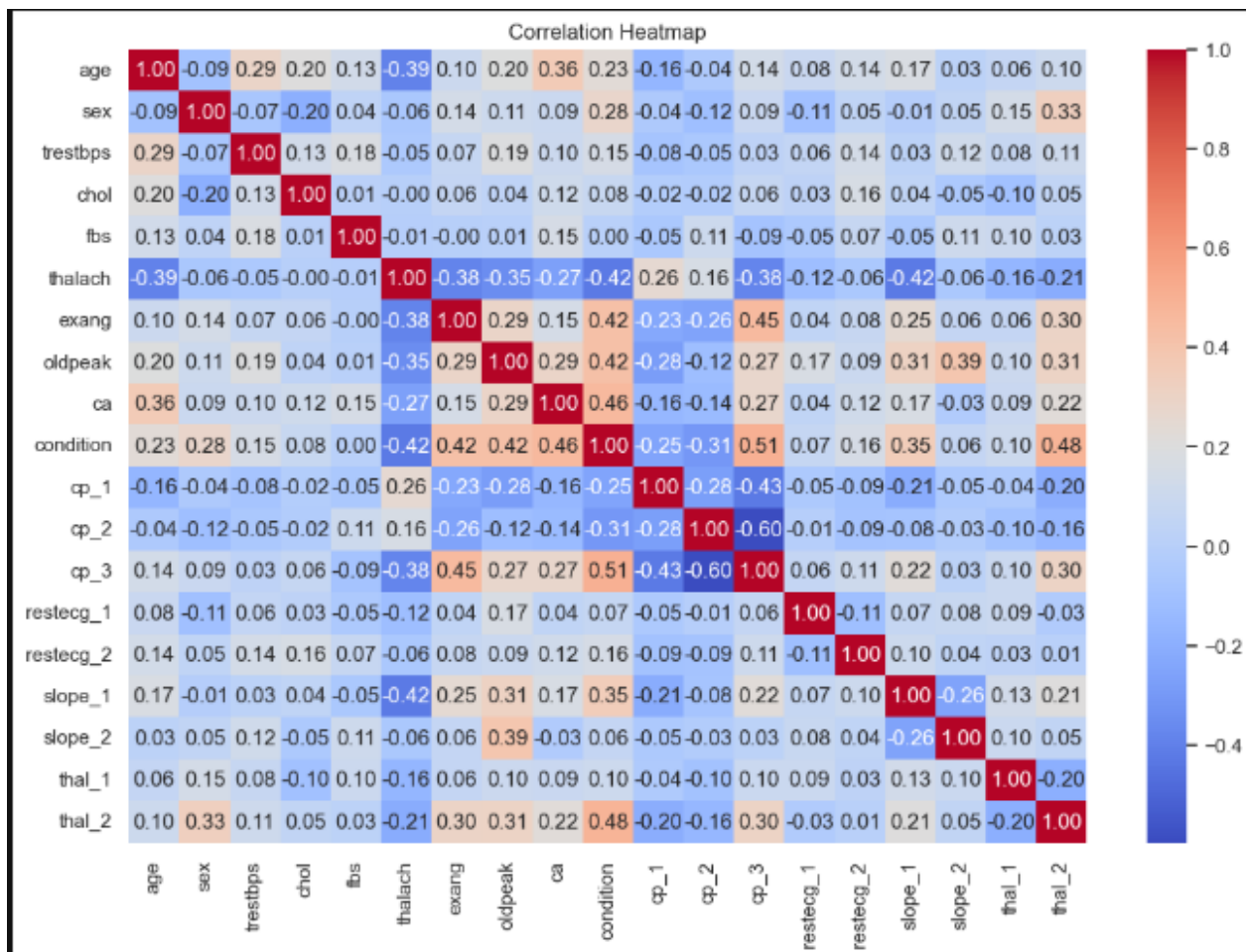
2-Target class distribution for balanced understanding.

Heart Disease Distribution
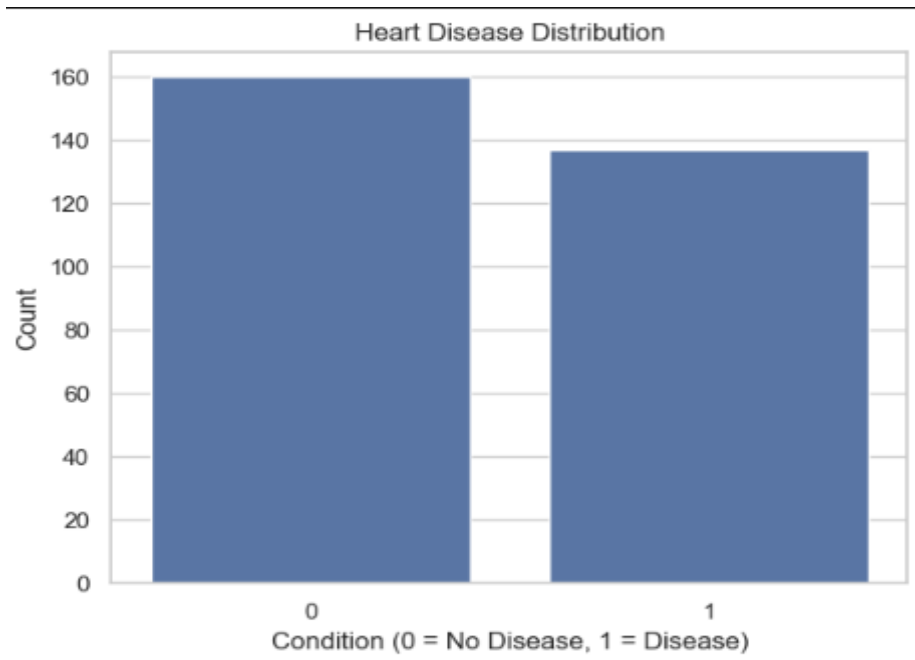
2. **Top features based on Random Forest importance.**

Top 10 Important Features from Random Forest

3. **Bar chart comparing F1 Scores of all models**

Model Comparison: F1 Score

These visualizations strengthen the understanding of what influences heart disease prediction.