

app

Get Code

Hyperparameter Tuning

Controls

# 🤖 AutoML — EDA + Clean + SHAP + Robust Engines (Option A)

Upload CSV dataset



Drag and drop file here

Limit 200MB per file • CSV

Browse files



heart.csv 11.1KB



✅ Dataset loaded

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope
0	63	1	3	145	233	1	0	150	0	2.3	
1	37	1	2	130	250	0	1	187	0	3.5	
2	41	0	1	130	204	0	0	172	0	1.4	
3	56	1	1	120	236	0	1	178	0	0.8	
4	57	0	0	120	354	0	1	163	1	0.6	

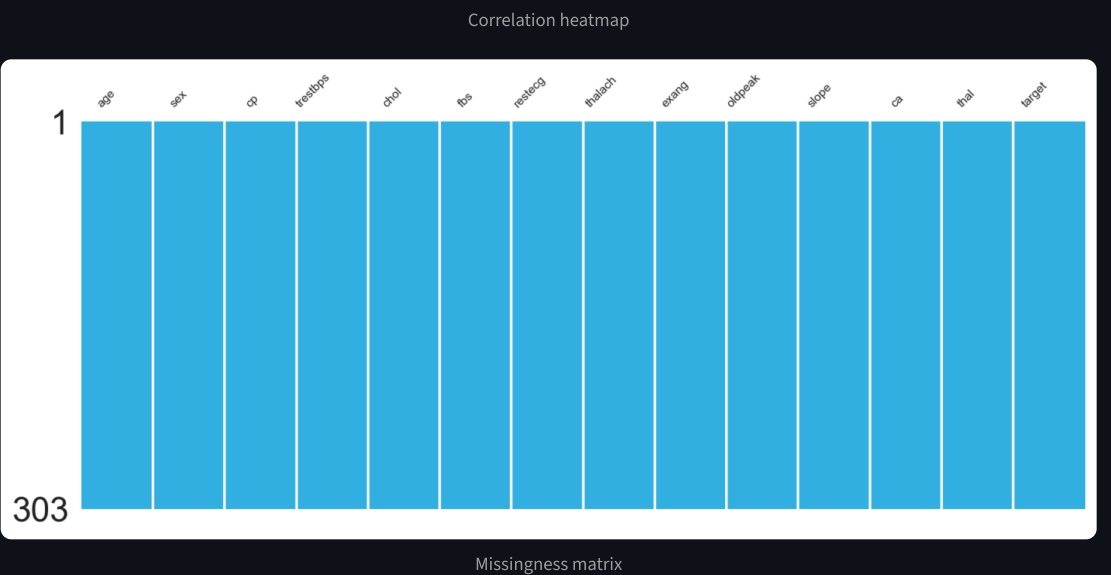
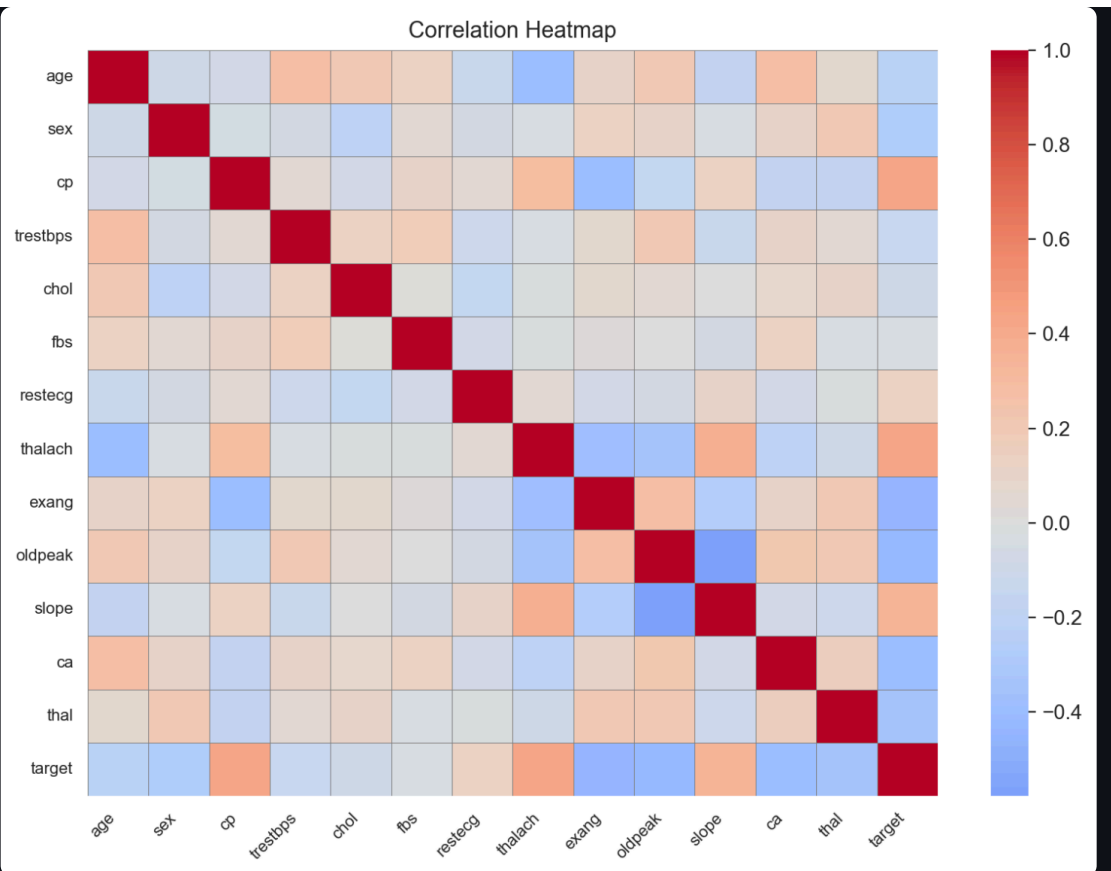


## Quick EDA Summary

- Rows: 303, Columns: 14

## Missing Values (top)

	dtype	missing_count	missing_pct
missing_count	unknown	0	
missing_pct	unknown	0	



## 2. Target, Task & Auto-clean

### Target & Task Type

Select target column

target

Task Type

Binary Classification

## Auto-clean & options

☒ Auto-clean (drop high-missing cols + impute)

Drop columns with missing ratio above:

0.50

Numeric fill strategy

mean






Categorical fill strategy

most\_frequent



## Feature Engineering Options

- ☒  Fix rare categories (< 5% freq)
- ☒  Handle outliers (IQR clipping)
- ☐  Create numeric interaction features
- ☐ Apply SMOTE on training set (if needed)

## Data issues detected

- Target cardinality: 2
- 1 duplicate rows found

## Detailed Auto-Clean Summary

Rows: 303 → 303 (0 rows dropped)


Columns: 14 → 14 (0 columns dropped)

No rows dropped due to missing target.

No columns dropped due to missing-value threshold.


Removed duplicate rows: 1

All missing values handled successfully.

 Feature Engineering applied successfully!

## Cleaned dataset summary


- Rows: 303, Columns: 14 (no detailed before/after info from cleaner).

 Train models (cleaned dataset)

### 3. Model Evaluation & Explainability

Choose model

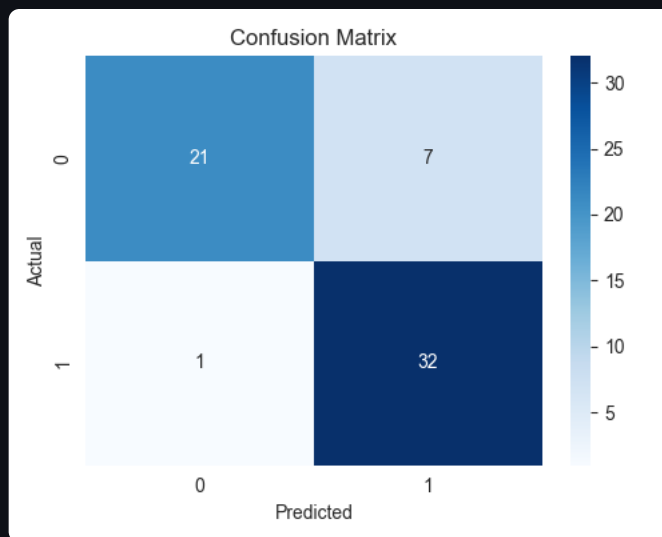
GaussianNB

 Get Code for This Model



#### Classification Metrics (on TEST set)

- Accuracy: 0.8689
- Precision (weighted): 0.8820
- Recall (weighted): 0.8689
- F1 (weighted): 0.8664



ROC/PR not available: Error opening 'reports/GaussianNB\_roc\_test.png'

Feature importance not available for this model.



#### Leaderboard (test scores)

Could not render leaderboard (mixed types?)



#### Manual prediction

age

42

sex

1

cp

3

trestbps

145

chol

233

fbs

1

restecg

0

thalach

150

exang

0

oldpeak

2.3

slope

0

ca

0

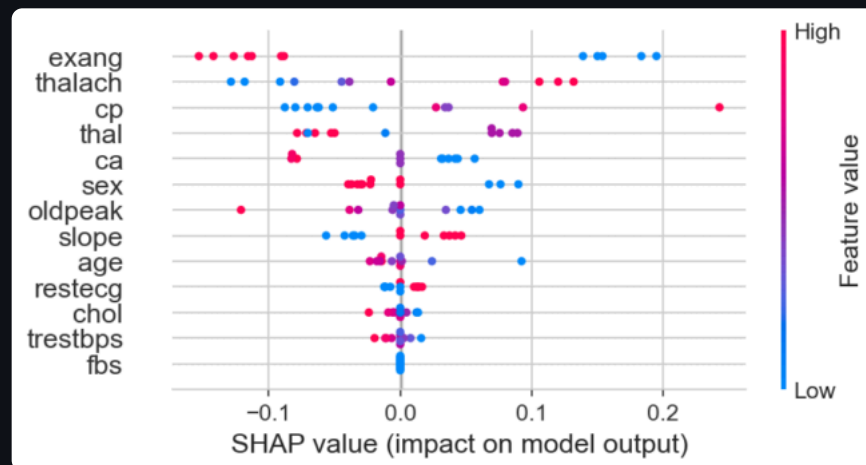
thal

1

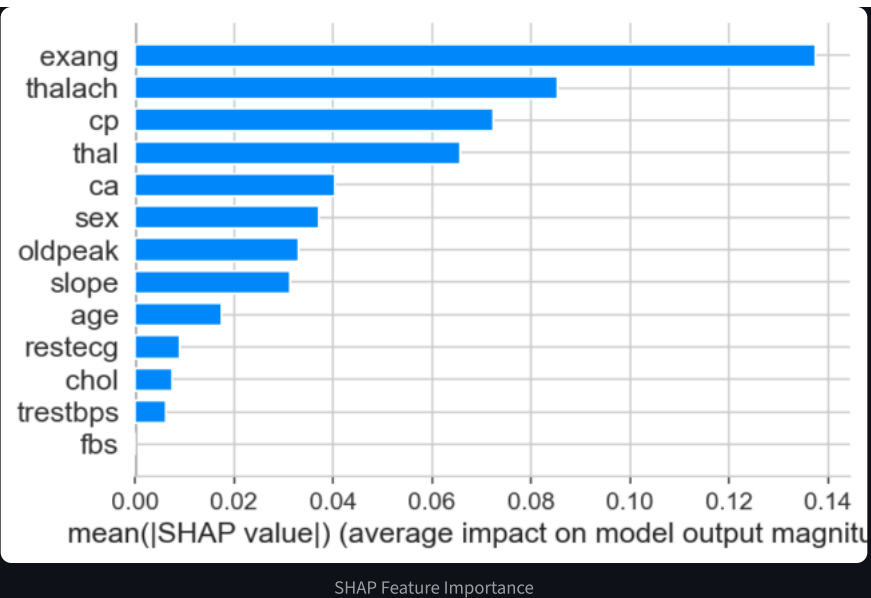
Predict

## SHAP Explainability (sample from TEST set)

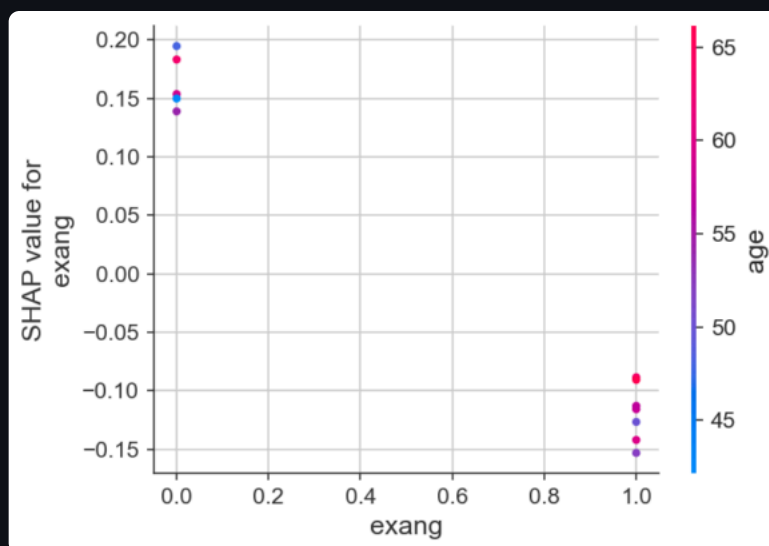
Compute SHAP (sampled, test set)



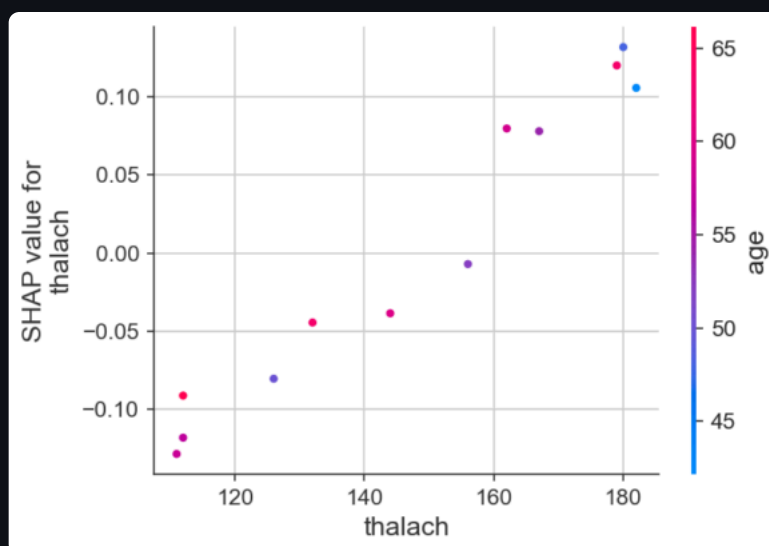
SHAP Summary



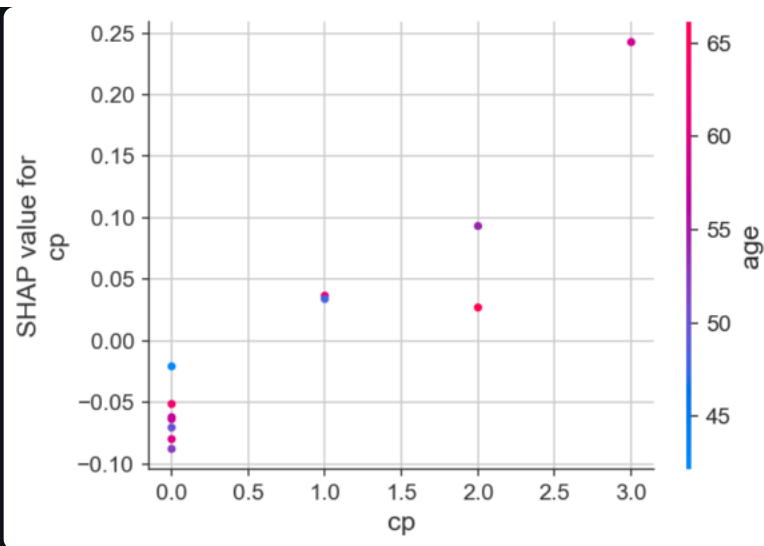
SHAP Feature Importance



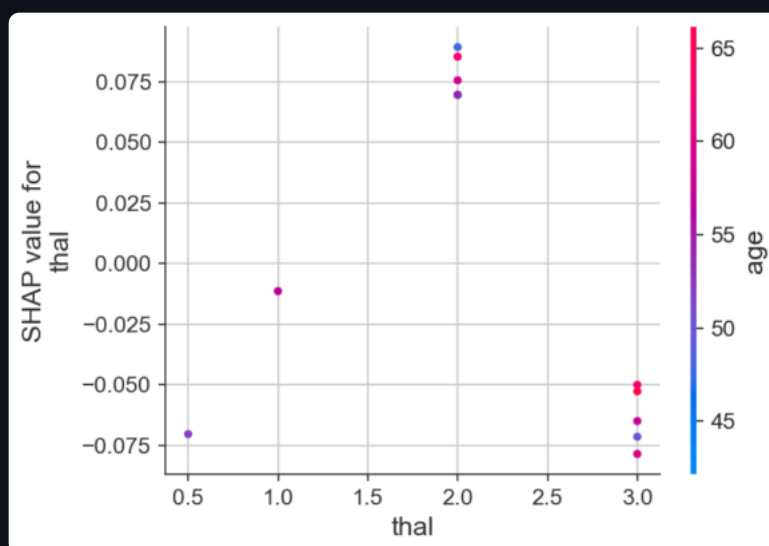
SHAP Dependence



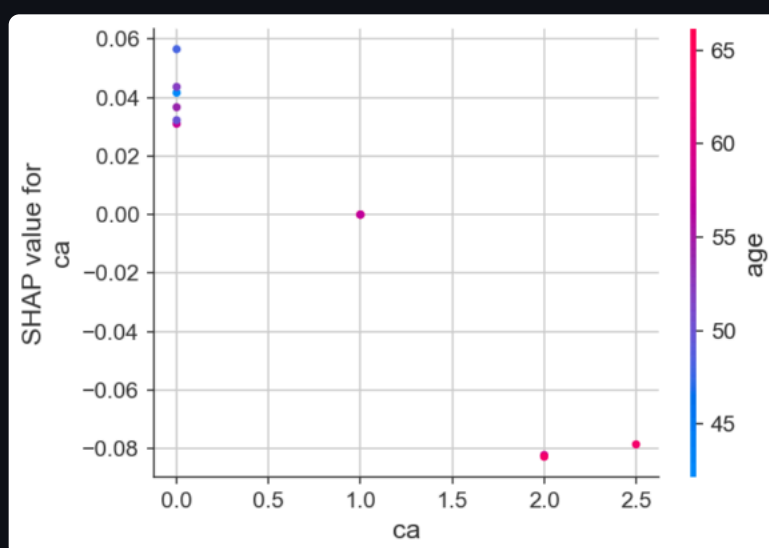
SHAP Dependence



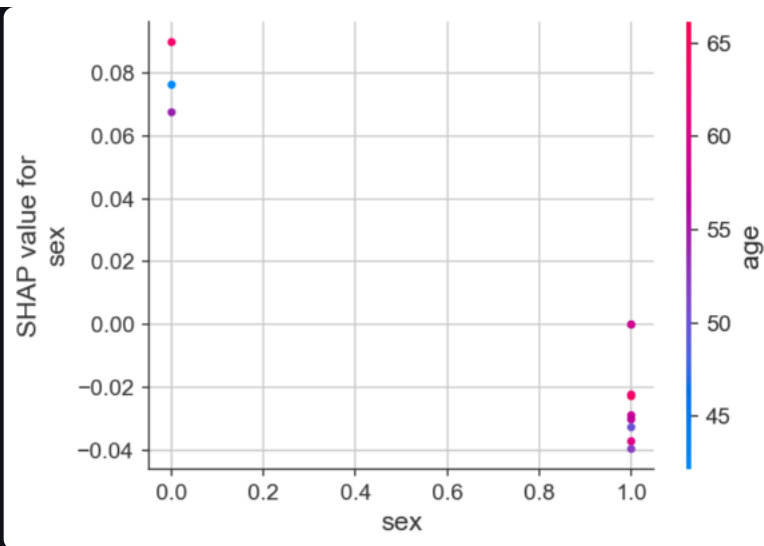
SHAP Dependence



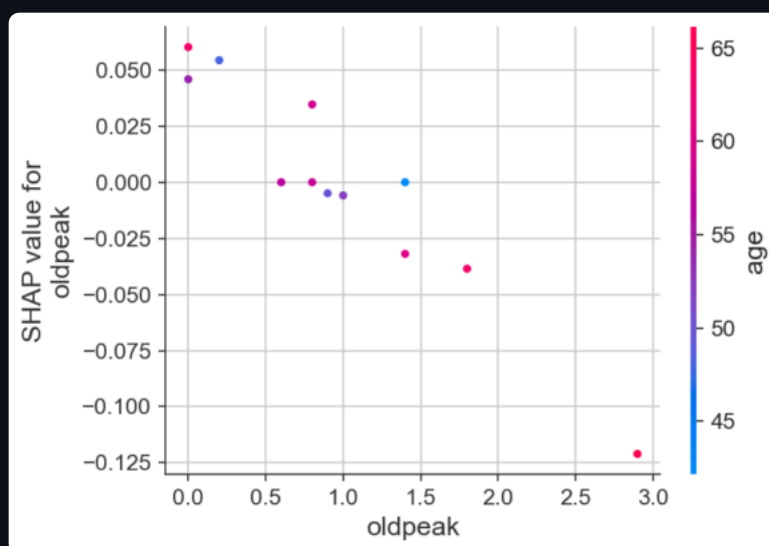
SHAP Dependence



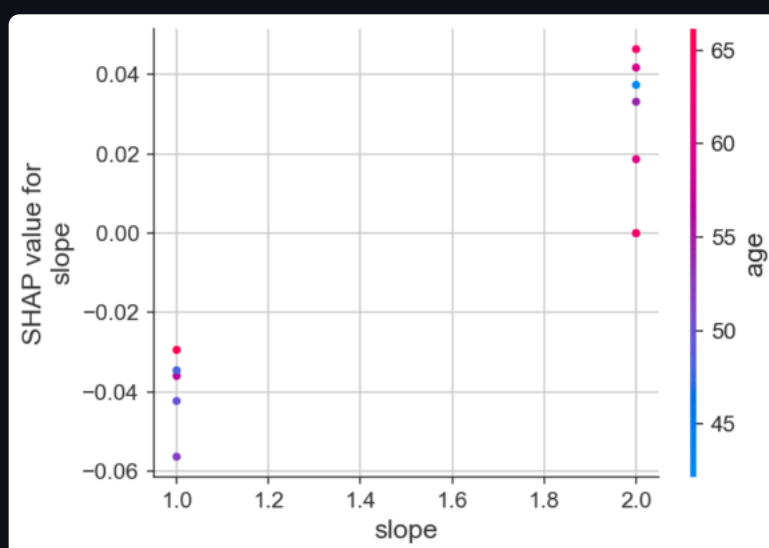
SHAP Dependence



SHAP Dependence

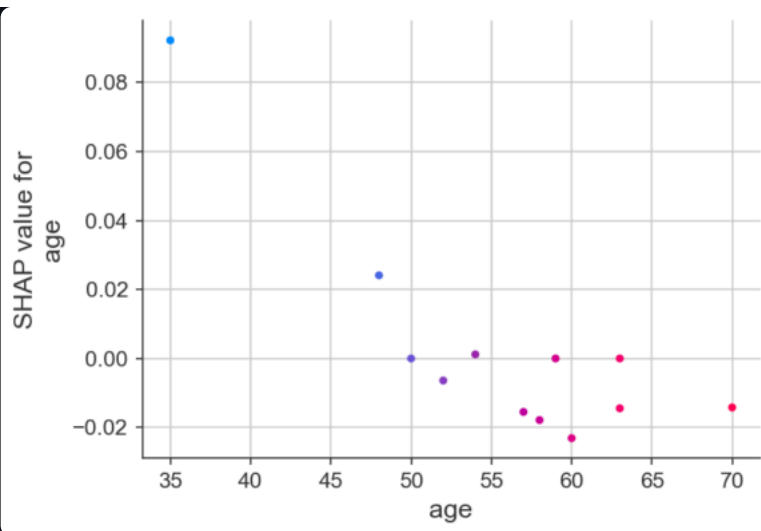


SHAP Dependence

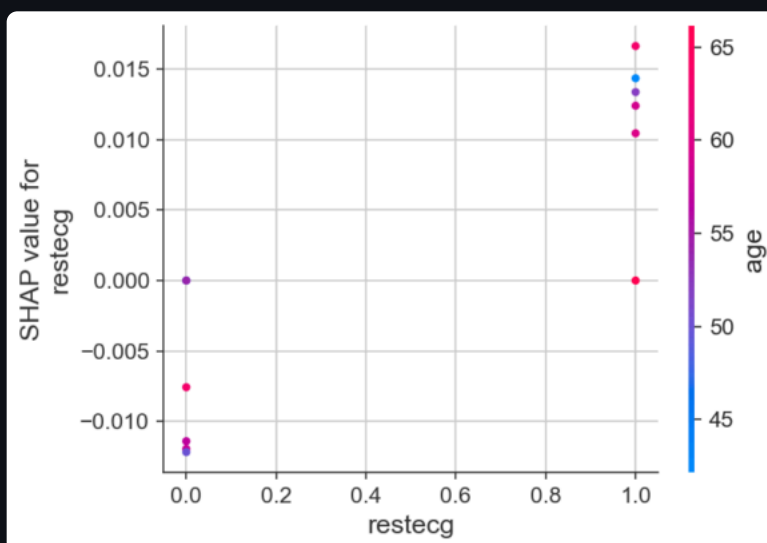


SHAP Dependence

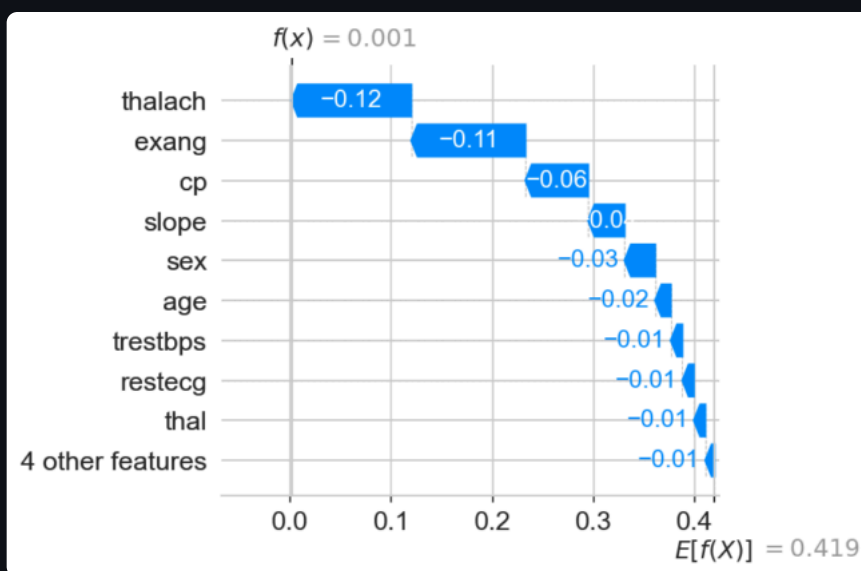




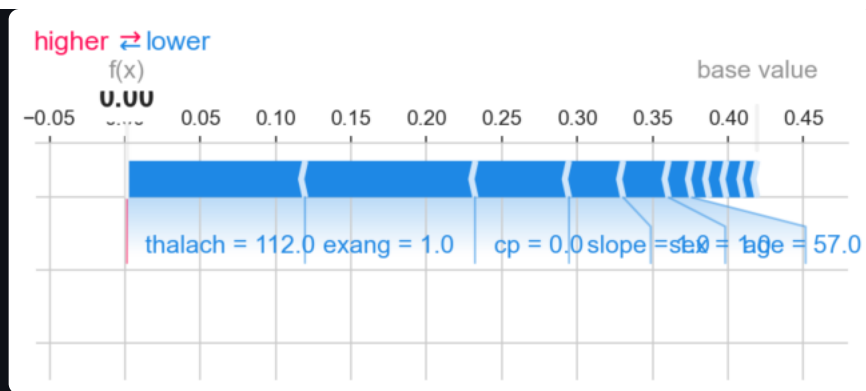
SHAP Dependence



SHAP Dependence



SHAP Waterfall



SHAP Force Plot

## SHAP Interactive Force Plot



### Export PDF Report

Generate PDF Report