

AutoML Final Report

AutoML Final Report

Generated: 2025-12-20 11:25:06

1. Dataset Overview

Original Dataset:

- Rows: 1025
- Columns: 14
- Missing Values: 0 (0.00%)
- Duplicate Rows: 723

Column Summary:

- Numeric Columns (14): age, sex, cp, trestbps, chol, fbs, restecg, thalach, exang, oldpeak, slope, ca, thal, target
- Categorical Columns (0):

After Preprocessing:

- Rows: 302
- Columns: 14
- Rows Removed: 723

2. EDA Findings

Numeric Features Statistics:

| Feature | Mean | Median | Std | Min | Max |
|----------|--------|--------|-------|--------|--------|
| age | 54.43 | 56.00 | 9.07 | 29.00 | 77.00 |
| sex | 0.70 | 1.00 | 0.46 | 0.00 | 1.00 |
| cp | 0.94 | 1.00 | 1.03 | 0.00 | 3.00 |
| trestbps | 131.61 | 130.00 | 17.52 | 94.00 | 200.00 |
| chol | 246.00 | 240.00 | 51.59 | 126.00 | 564.00 |
| fbs | 0.15 | 0.00 | 0.36 | 0.00 | 1.00 |

| | | | | | |
|---------|--------|--------|-------|-------|--------|
| restecg | 0.53 | 1.00 | 0.53 | 0.00 | 2.00 |
| thalach | 149.11 | 152.00 | 23.01 | 71.00 | 202.00 |
| exang | 0.34 | 0.00 | 0.47 | 0.00 | 1.00 |
| oldpeak | 1.07 | 0.80 | 1.18 | 0.00 | 6.20 |
| slope | 1.39 | 1.00 | 0.62 | 0.00 | 2.00 |
| ca | 0.75 | 0.00 | 1.03 | 0.00 | 4.00 |
| thal | 2.32 | 2.00 | 0.62 | 0.00 | 3.00 |
| target | 0.51 | 1.00 | 0.50 | 0.00 | 1.00 |

3. Data Quality Issues Detected

Issues Summary:

- High Severity: 3
- Medium Severity: 2
- Low Severity: 3

Issues Detected:

- [HIGH] DUPLICATE_ROWS in 'N/A': 723 (70.54%)
- [MEDIUM] OUTLIERS in 'trestbps': 30 (2.93%)
- [MEDIUM] OUTLIERS in 'chol': 16 (1.56%)
- [HIGH] OUTLIERS in 'fbs': 153 (14.93%)
- [LOW] OUTLIERS in 'thalach': 4 (0.39%)
- [LOW] OUTLIERS in 'oldpeak': 7 (0.68%)
- [HIGH] OUTLIERS in 'ca': 87 (8.49%)
- [LOW] OUTLIERS in 'thal': 7 (0.68%)

4. Preprocessing Decisions

Methods Applied:

- Missing Values Strategy: Median
- Outlier Handling: Remove
- Scaling Method: Standard
- Encoding Method: OneHot
- Test Size: 0.2

Impact Summary:

- Rows Removed: 723
- Missing Values Reduced: 0 → 0

- Features Modified: 14 (from 14)

5. Model Configurations & Hyperparameters

Training Configuration:

- Test Size: 0.2
- Random State: 42
- Hyperparameter Tuning: Yes

Models Trained: 14 total

Hyperparameter Tuned Models (7):

• Logistic Regression (Tuned)

- C: 1
- max_iter: 100
- penalty: l2
- solver: lbfgs

• K-Neighbors Classifier (Tuned)

- metric: manhattan
- n_neighbors: 5
- weights: distance

• Decision Tree Classifier (Tuned)

- criterion: gini
- max_depth: 10
- minsamplesleaf: 1
- minsamplessplit: 2

• Gaussian Naive Bayes (Tuned)

- var_smoothing: 0.012915496650148827

• Random Forest (Tuned)

- criterion: gini
- max_depth: None
- minsamplesleaf: 1
- minsamplessplit: 2
- n_estimators: 50

• Support Vector Machine (Tuned)

- C: 100

- gamma: scale
- kernel: rbf

- **Decision Tree Rule-based (Tuned)**

- criterion: gini
- max_depth: 10
- minsamplesleaf: 1
- minsamplessplit: 2

Non-Tuned Models (7):

- Logistic Regression (default parameters)
- K-Neighbors Classifier (default parameters)
- Decision Tree Classifier (default parameters)
- Gaussian Naive Bayes (default parameters)
- Random Forest (default parameters)
- Support Vector Machine (default parameters)
- Decision Tree Rule-based (default parameters)

6. Model Performance Comparison

| Model | Accuracy | Precision | Recall | F1-Score | Training Time |
|----------------------------------|----------|-----------|--------|----------|---------------|
| Logistic Regression | 0.7951 | 0.8023 | 0.7951 | 0.7938 | 0.01s |
| K-Neighbors Classifier | 0.8341 | 0.8387 | 0.8341 | 0.8335 | 0.01s |
| Decision Tree Classifier | 0.9854 | 0.9858 | 0.9854 | 0.9854 | 0.01s |
| Gaussian Naive Bayes | 0.8000 | 0.8105 | 0.8000 | 0.7982 | 0.00s |
| Random Forest | 0.9854 | 0.9858 | 0.9854 | 0.9854 | 0.36s |
| Support Vector Machine | 0.8878 | 0.8923 | 0.8878 | 0.8875 | 0.23s |
| Decision Tree Rule-based | 0.8439 | 0.8604 | 0.8439 | 0.8420 | 0.01s |
| Logistic Regression (Tuned) | 0.7951 | 0.8023 | 0.7951 | 0.7938 | 26.84s |
| K-Neighbors Classifier (Tuned) | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.51s |
| Decision Tree Classifier (Tuned) | 0.9854 | 0.9858 | 0.9854 | 0.9854 | 2.32s |
| Gaussian Naive Bayes (Tuned) | 0.8000 | 0.8105 | 0.8000 | 0.7982 | 0.28s |
| Random Forest (Tuned) | 0.9854 | 0.9858 | 0.9854 | 0.9854 | 153.91s |
| Support Vector Machine (Tuned) | 0.9854 | 0.9858 | 0.9854 | 0.9854 | 18.65s |
| Decision Tree Rule-based (Tuned) | 0.9854 | 0.9858 | 0.9854 | 0.9854 | 2.17s |

7. Best Model Summary & Justification

Selected Model: K-Neighbors Classifier (Tuned)

Reason: Best F1 score: 1.0000

Performance Metrics:

- Accuracy: 1.0
- Precision: 1.0
- Recall: 1.0
- F1-Score: 1.0
- ROC-AUC: 1.0
- Training Time: 1.5126826763153076s

Hyperparameters:

- metric: manhattan
- n_neighbors: 5
- weights: distance