FAIKR mod.3

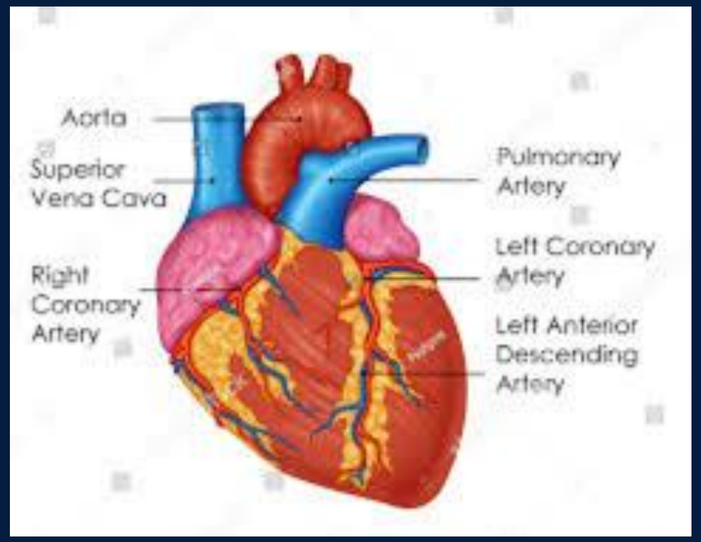
ARTIFICIAL INTELLIGENCE

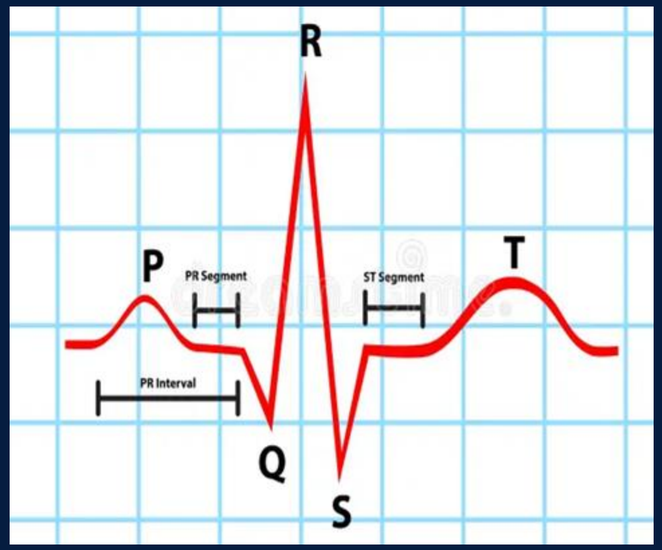
BAYESIAN HEART FAILURE PREDICTION

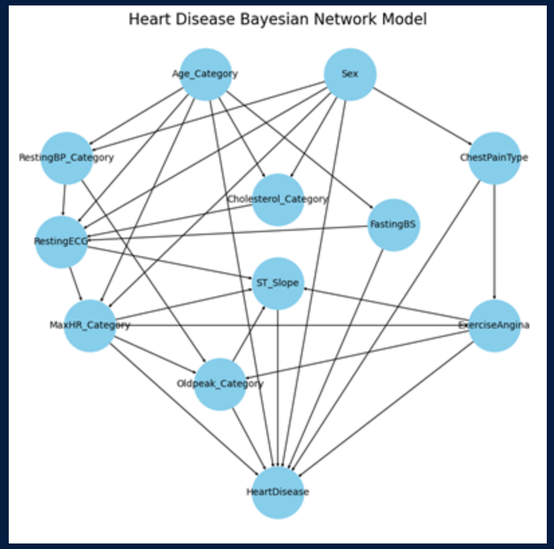
Presented by: Mauro Dore, Leonardo Monti

INTRODUCTION

THE PROJECT: DIAGNOSING AND ASSESSING THE RISK OF HEART DISEASES USING BAYESIAN NETWORKS MODELS

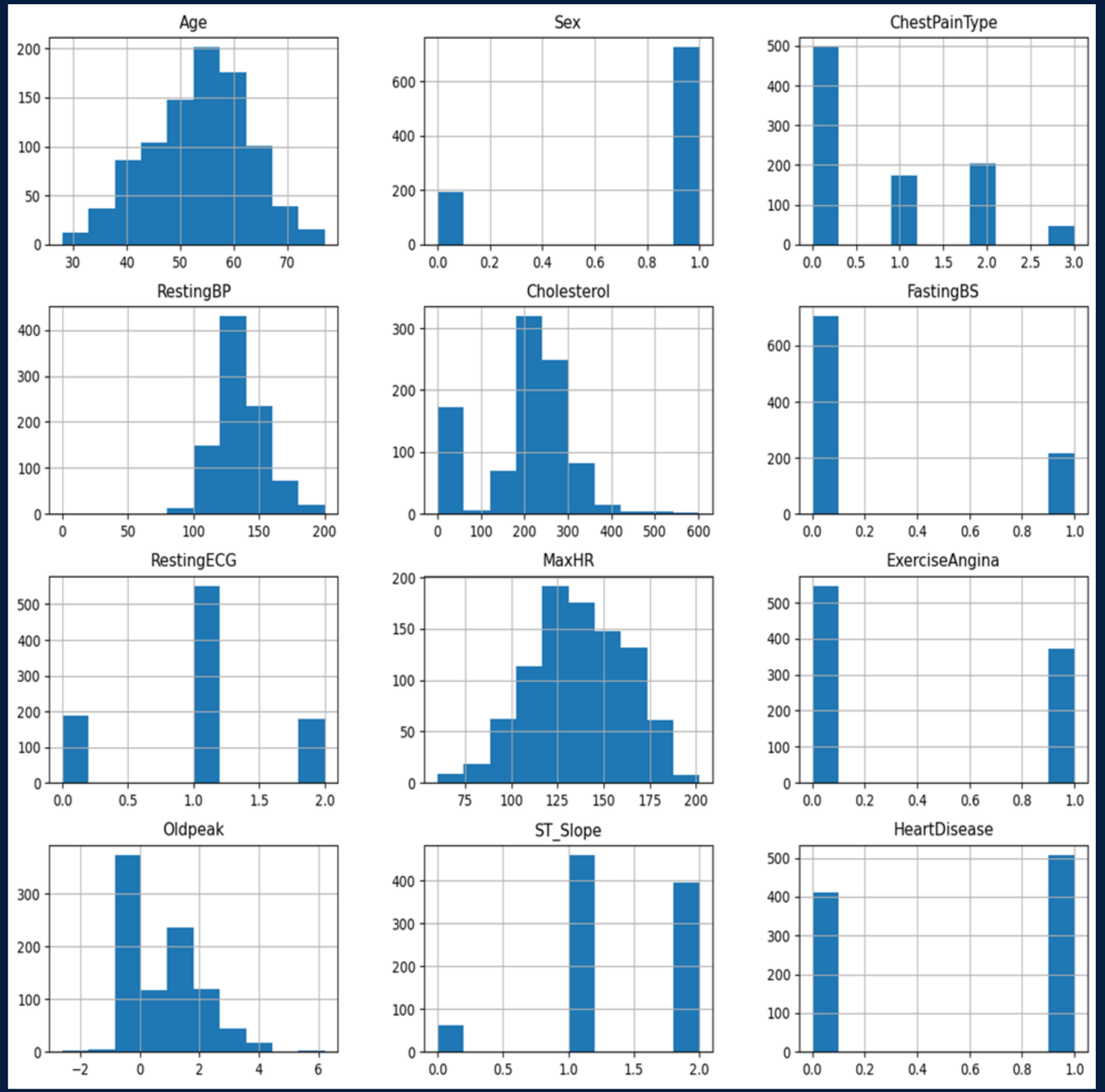






THE DATA

- DEMOGRAPHIC INFORMATIONS
- CLINICAL MEASUREMENTS
- UNDER STRESS RESULTS



Distribution of the data

PREPROCESSING

Encoding

Categorical features were converted into numerical representations.

Binning

Numerical data were binned into categories

Imputation

Values of 0 (e.g. Cholesterol) were identified as "nulls in disguise" and were imputed with the median.

Check

Duplicate rows were checked and any missing value.

THE NETWORKS

CUSTOM MODEL

- Includes all the features
- Casual links designed following correlation and authors' knowledge

SEARCH BASED MODELS

- Tree model
- HillClimb model
- Constraint-Based

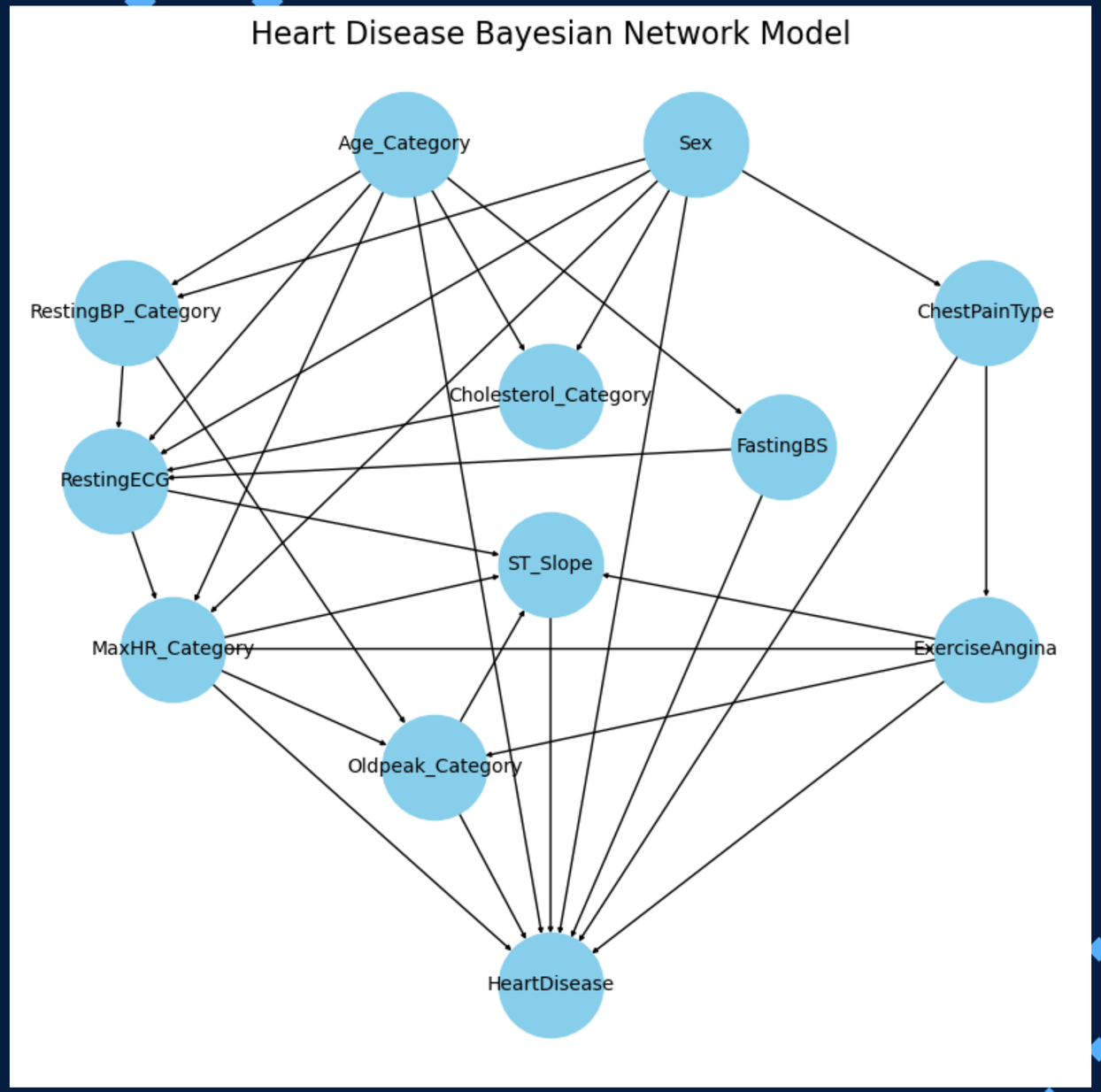
Causal links learned automatically from the dataset

CUSTOM MODEL

- Some nodes must be root (age)
- Almost everything influences Heart Disease
- Age and heart rate are directly correlated
- Chest pain influences exercise angina

Diagnosis performance:

```
Accuracy on the test set: 0.8478
F1 on the test set: 0.8654
```

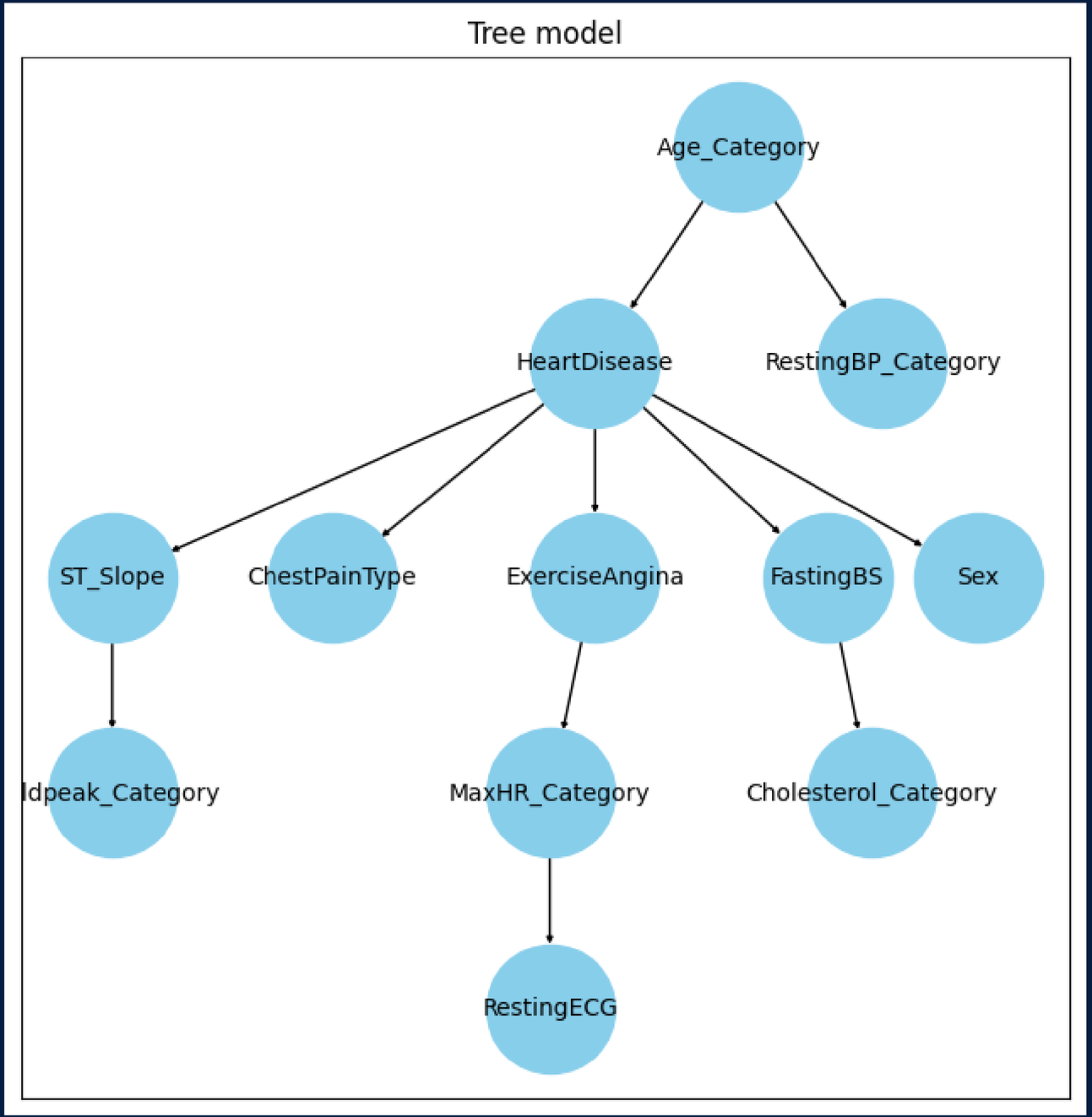


TREE MODEL

- Learned automatically from the dataset
- T**ree search algorithm**
- No additional constraints.
- The structure is more simple
- Heart disease explains many symptoms

Diagnosis performance:

```
Accuracy on the test set: 0.8533
F1 on the test set: 0.8720
```

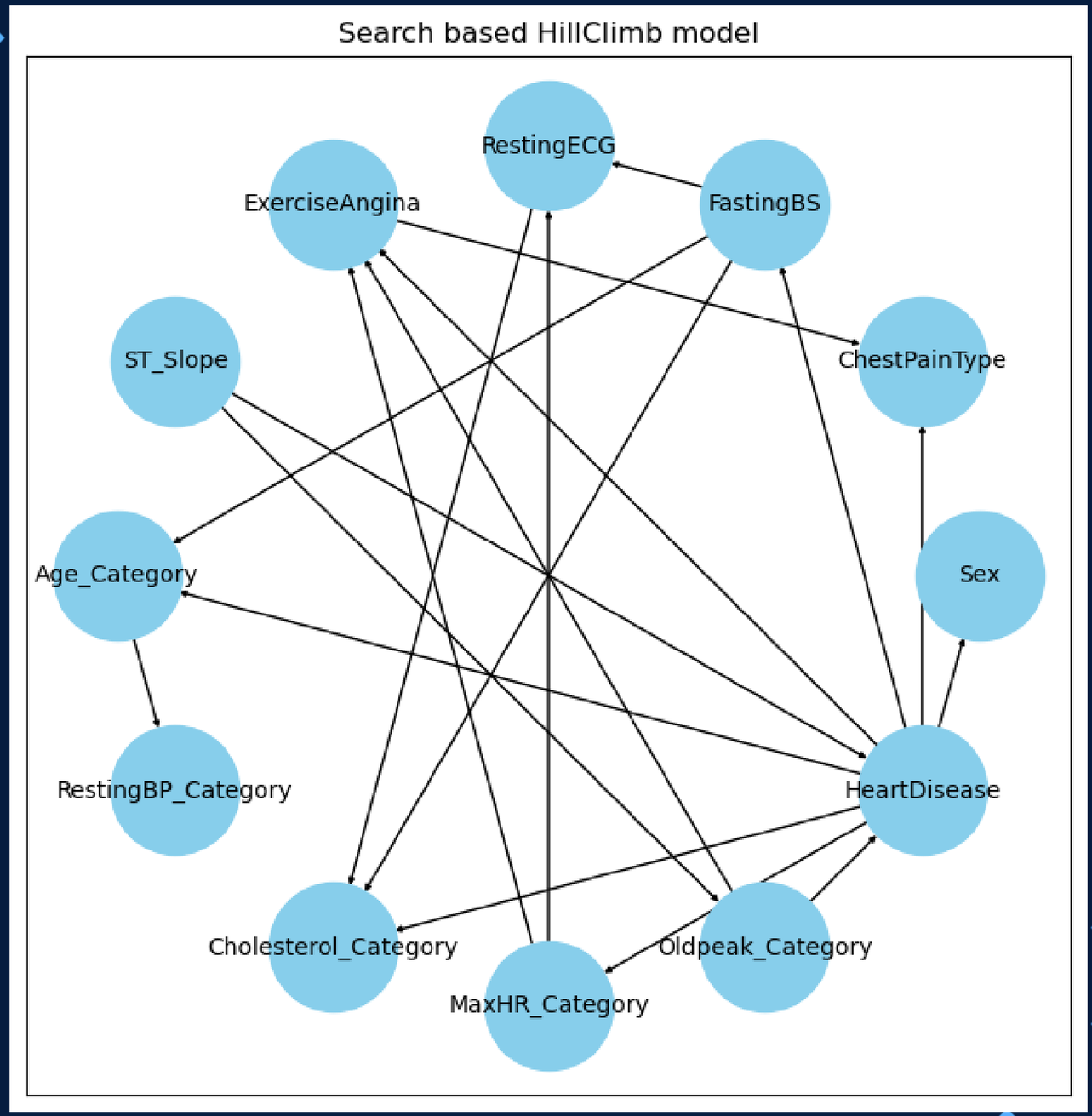


HILL – CLIMB MODEL

- Learned automatically from the dataset
- **Hill-climb search** algorithm
- No additional constraints
- **BDeu score** as scoring function
- The result has more connections
- Heart disease is the node with the most links

Diagnosis performance:

```
Accuracy on the test set: 0.8750
F1 on the test set: 0.8959
```

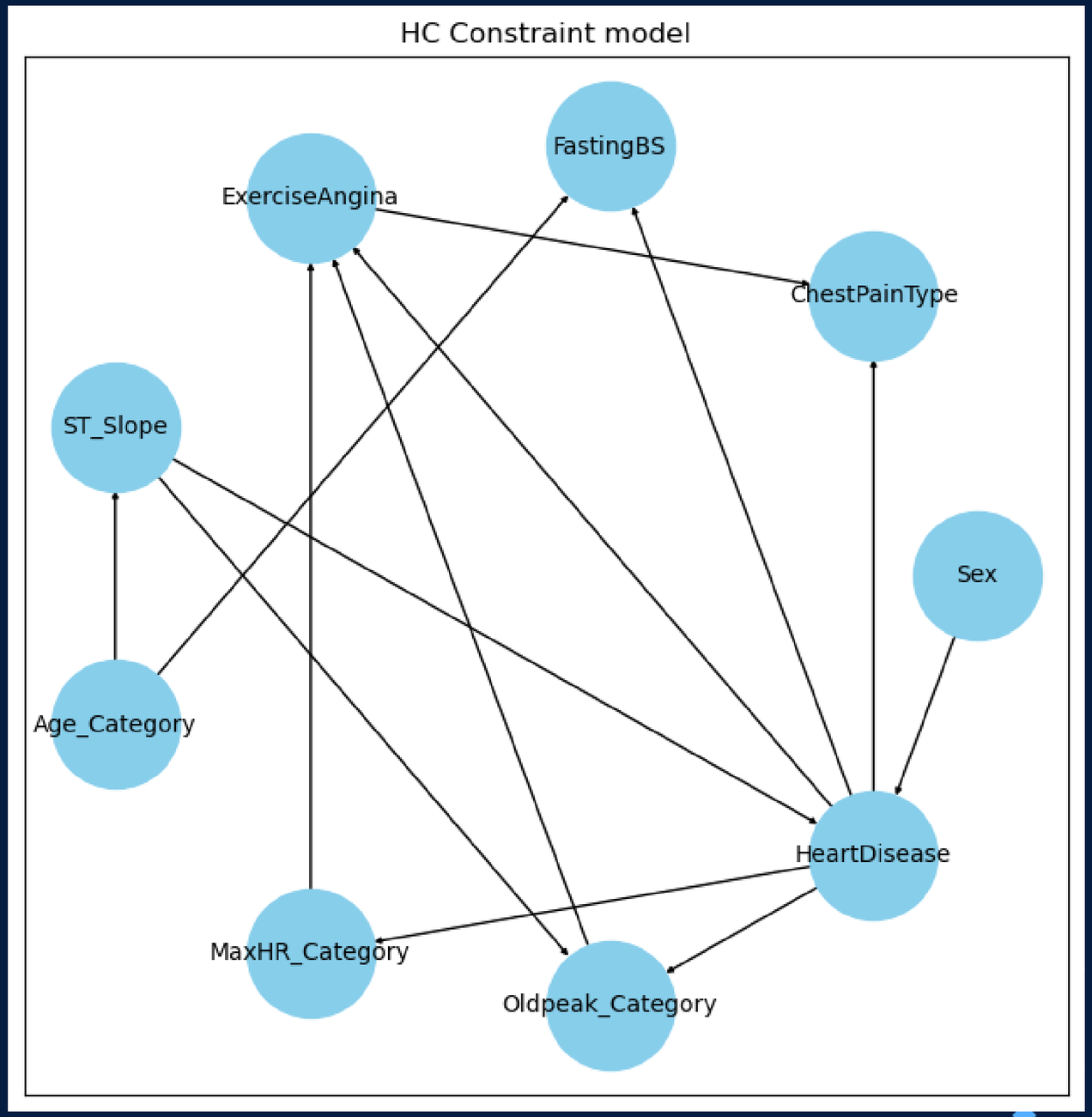


CONSTRAINT MODEL

- Learned automatically from the dataset
- **Hill-climb search** algorithm
- Constraints from the study of the previous models
- Pruned the nodes with small correlation or not in the Markov Blankets.

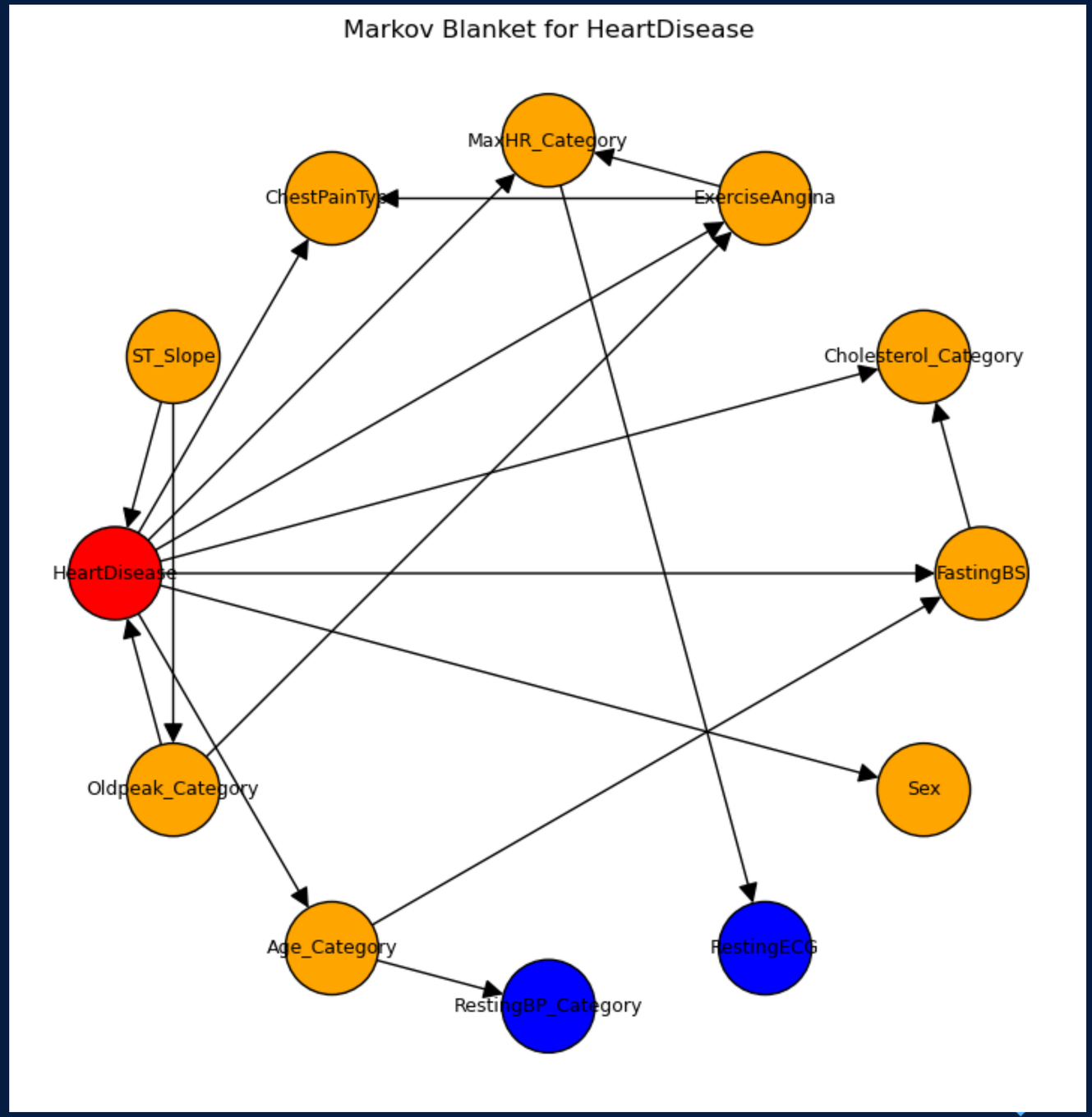
Diagnosis performance:

```
Accuracy on the test set: 0.8696
F1 on the test set: 0.8909
```



MARKOV BLANKET

- Focus on Heart Disease
- Nearly all features in the MB
- Used to prune nodes in custom model

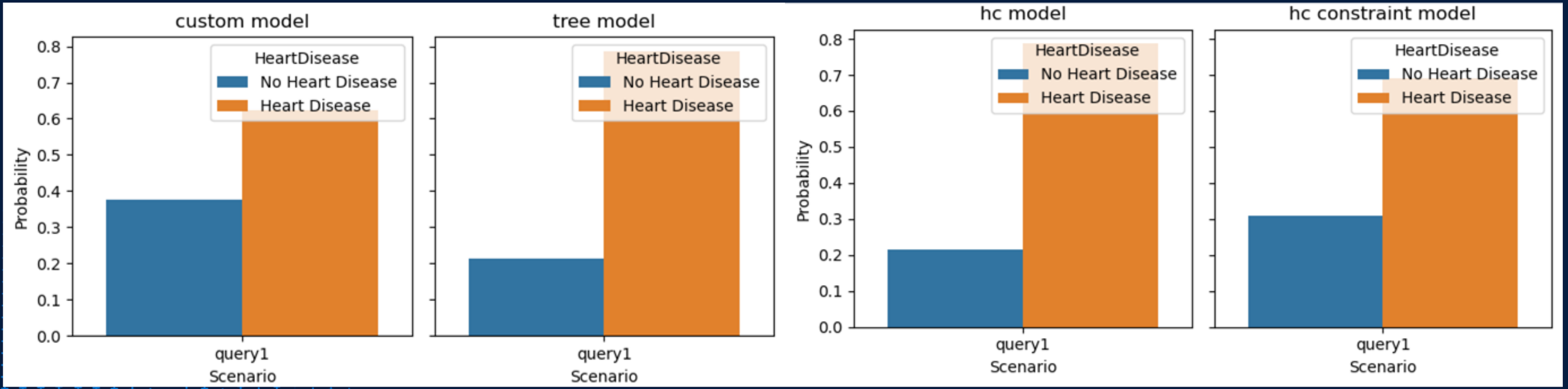


MB for Hill Climb model

QUERY 1

What is the risk of Heart Disease for old male patients

$$P(HeartDisease \mid Age = 2, Sex = 1)$$



QUERY 2

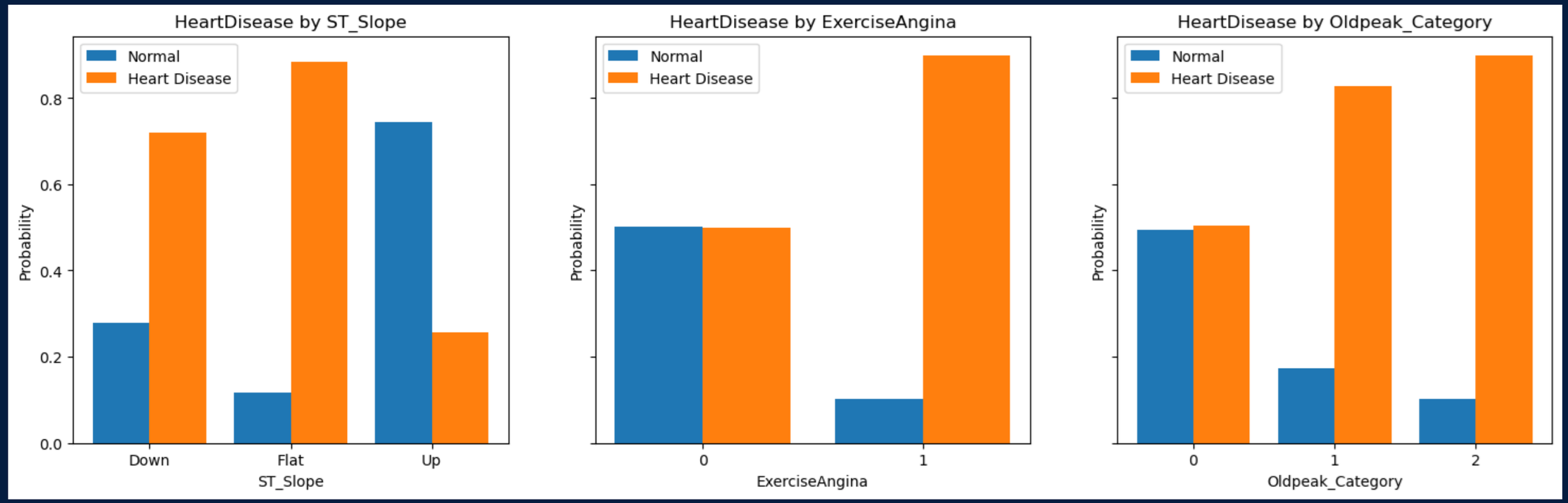
Which activity-related features cause a high risk of heart disease in an old male

$$P(HeartDisease \mid Age = 2, Sex = 1, ST_{Slope} = 1)$$

$$P(HeartDisease \mid Age = 2, Sex = 1, ExerciseAngina = 1)$$

$$P(HeartDisease \mid Age = 2, Sex = 1, Oldpeak = 1)$$

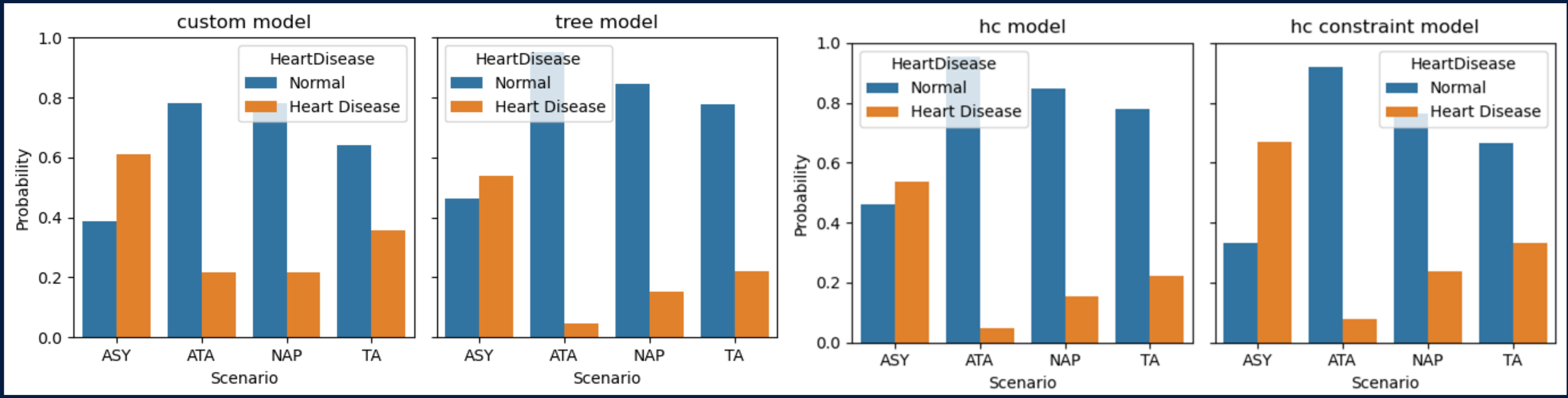
Results from Tree Model



QUERY 3

Heart disease risk by chest pain type, for young patients comparison among models

$$P(HeartDisease \mid Age = 0, ChestPainType = C)$$
$$with\ C\ \in \{0, 1, 2, 3\}$$



PERFORMANCE EVALUATION

- Task: query to diagnose Heart Disease from a subset of the features
- Average execution time over 1000 repetitions
- Very fast execution

```
Elapsed time for custom:  0.0005539545059204101 seconds
Elapsed time for tree:  0.0001422632932662964 seconds
Elapsed time for hc:  0.0002874422073364257 6 seconds
Elapsed time for hc constraint:  0.0002303251266479492 seconds
```

CONCLUSIONS

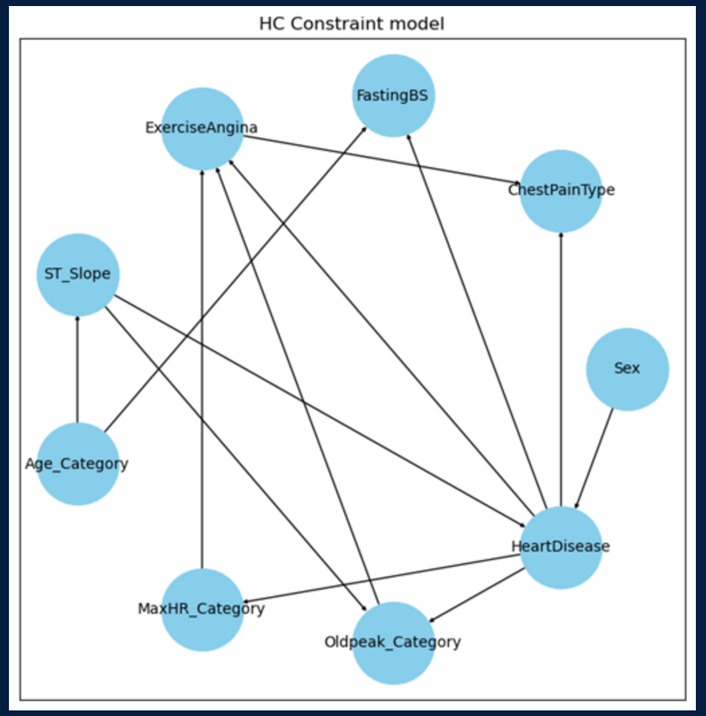
In this project we show an application of BN for the diagnosis and explanation of CVDs starting from clinical features.

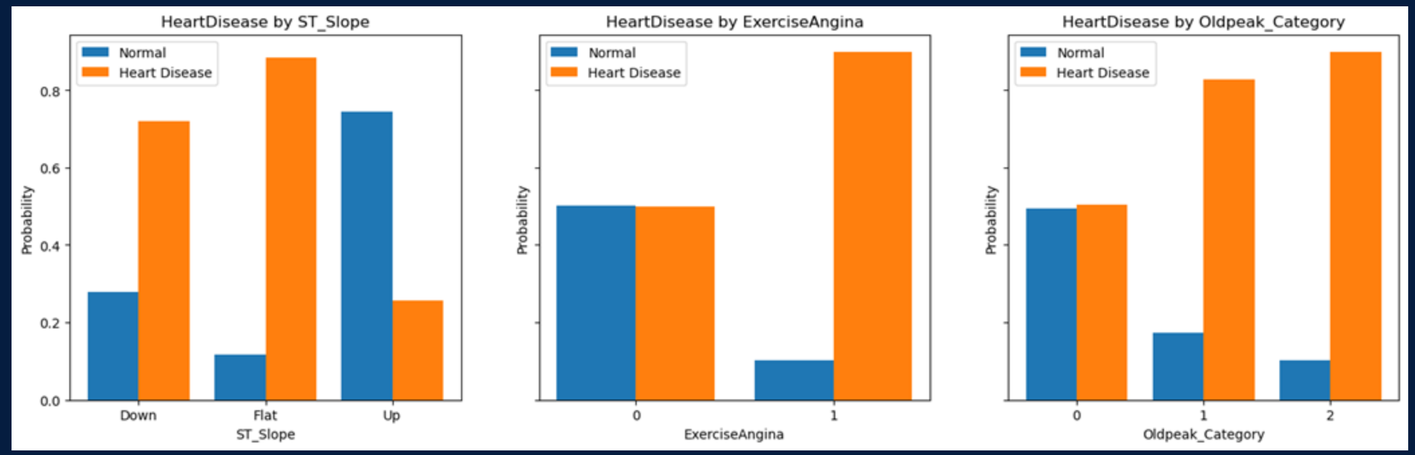
Best performance is obtained by the Hill Climb model with constraints.
This model is also the most efficient.

```
Evaluating model: hc
Accuracy on the test set: 0.8750
F1 on the test set: 0.8959
```

Seeing the results and the nature of the dataset, we advise to fetch data about more characteristics of the patients since the technique proved to be efficient. We also advise to seek guidance from medical experts in order to provide better constraints to the Hill Climb model since we see its potential improvement.

Thanks


HC Constraint model
FastingBS
ExerciseAngina
ChestPainType
ST_Slope
Sex
Age_Category
HeartDisease
MaxHR_Category
Oldpeak_Category


HeartDisease by ST_Slope
Normal
Heart Disease
Probability
0.0
0.2
0.4
0.6
0.8
Down
Flat
Up
ST_Slope
HeartDisease by ExerciseAngina
Normal
Heart Disease
Probability
0
1
ExerciseAngina
HeartDisease by Oldpeak_Category
Normal
Heart Disease
Probability
0
1
2
Oldpeak_Category