Cardiovascular Disease Analysis

Name: Dabir Aishwarya (19CO012)
Khedkar Pratiksha (19CO045)
Panch Laxmi (19CO053)
Thakare Tejal (19CO068)

Guide: Dr. M. A. Pradhan

Machine Learning

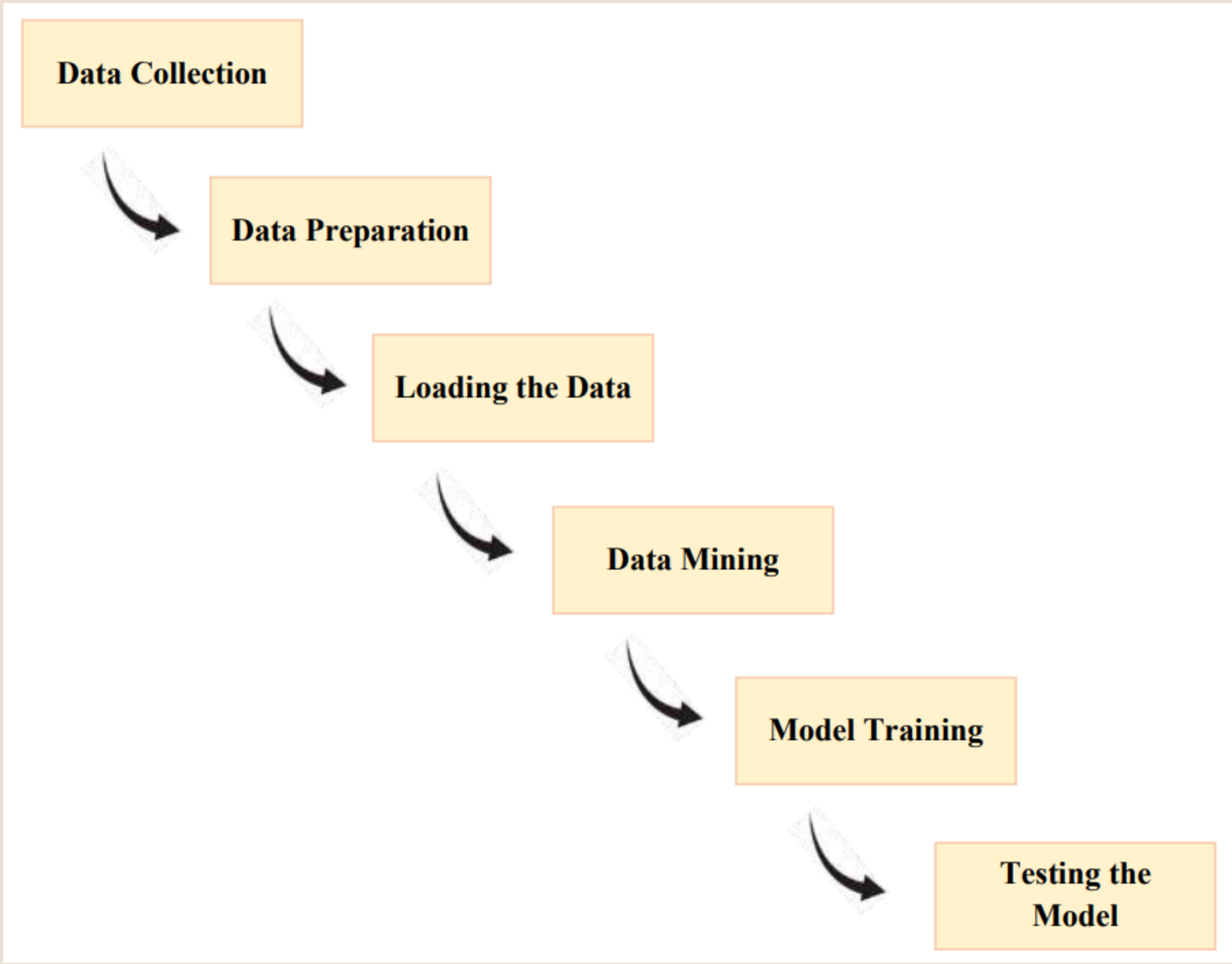
Problem Statement

- Cardiovascular disease is becoming a major cause of death all over the world.
- ML algorithms combined with new computing technologies promote scalability and improve efficiency.
- For providing visual representation to the doctor as we can’t take risk with the life we have the second opinion for the doctor with the parameters and visual representation which will help for better understanding of the doctor.
- Explanatoins generated by our model can give us visual representation and the most affecting factors

Motivation

- Early Detection of Diseases
- Logical Diagnosis
- Decision Tool for Medical Expert
- Correlation in Data
- Generate Explanations

Workflow


Data Collection
Data Preparation
Loading the Data
Data Mining
Model Training
Testing the Model

Dataset

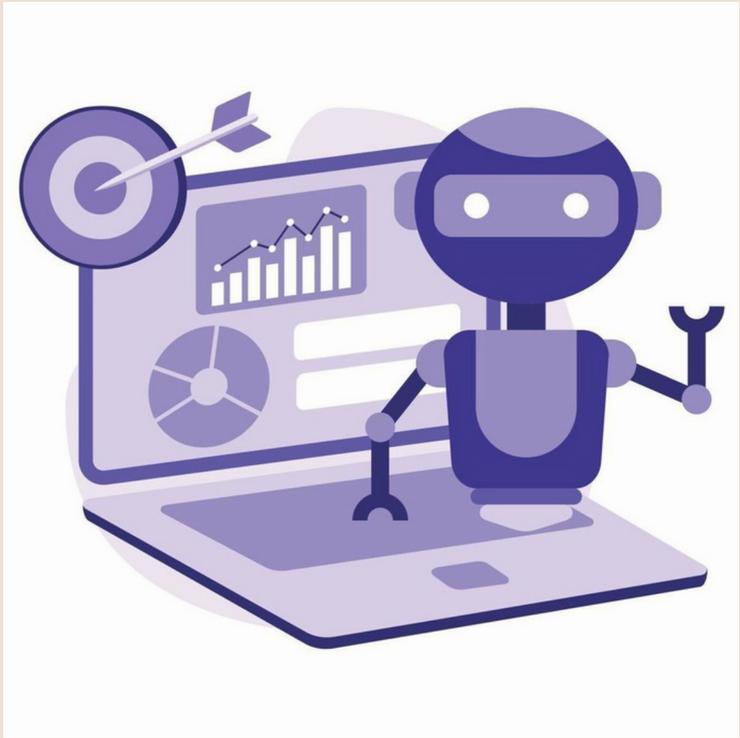
Dataset was taken from the UCI repository. It is Binary datset and output is 0 and 1.

- 0: No disease
- 1: Disease is present

	Age	Sex	ChestPainType	RestingBP	Cholesterol	FastingBS	RestingECG	MaxHR	ExerciseAngina	Oldpeak	ST_Slope	HeartDisease
0	40	M	ATA	140	289	0	Normal	172	N	0.0	Up	0
1	49	F	NAP	160	180	0	Normal	156	N	1.0	Flat	1
2	37	M	ATA	130	283	0	ST	98	N	0.0	Up	0
3	48	F	ASY	138	214	0	Normal	108	Y	1.5	Flat	1
4	54	M	NAP	150	195	0	Normal	122	N	0.0	Up	0

Steps

- Import Dataset
- Data preprocessing
- Splitting dataset
- Model training
- Generating explainations



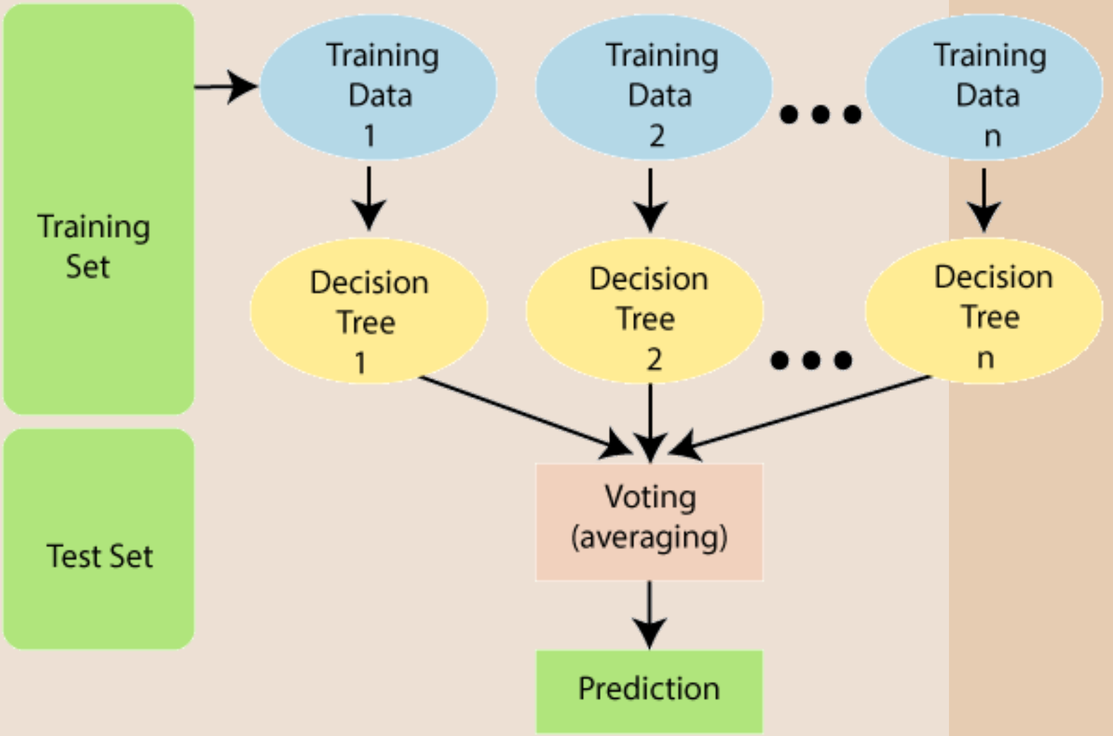
Data preprocessing

- Check for missing values
- Label encoding
- Normalization

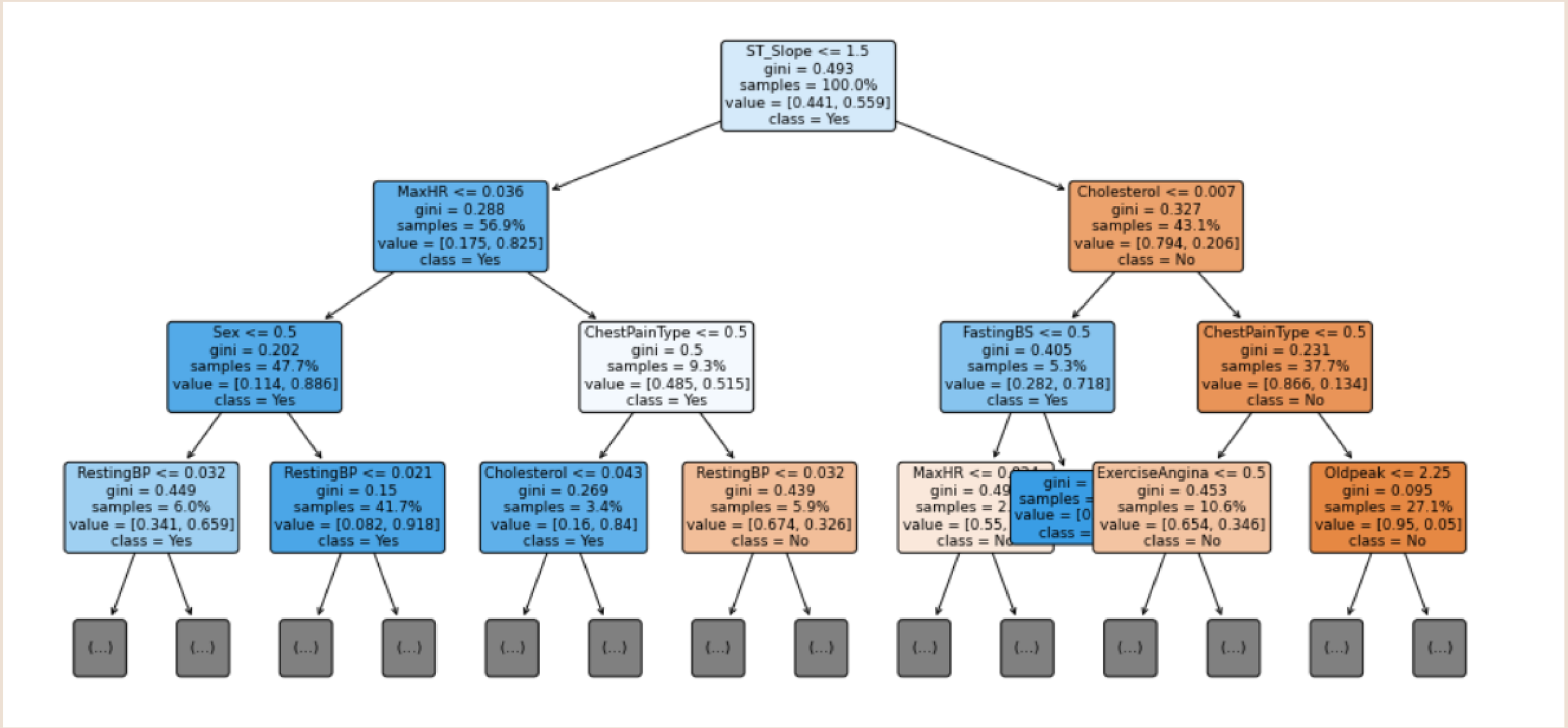
Splitting the dataset

- Train test split and model training
- K Fold validation

Random Forest



- Random forests or random decision forests is an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time
- Gini index is used as measure of impurity for random forest.
- The Gini Index is determined by deducting the sum of squared of probabilities of each class from one

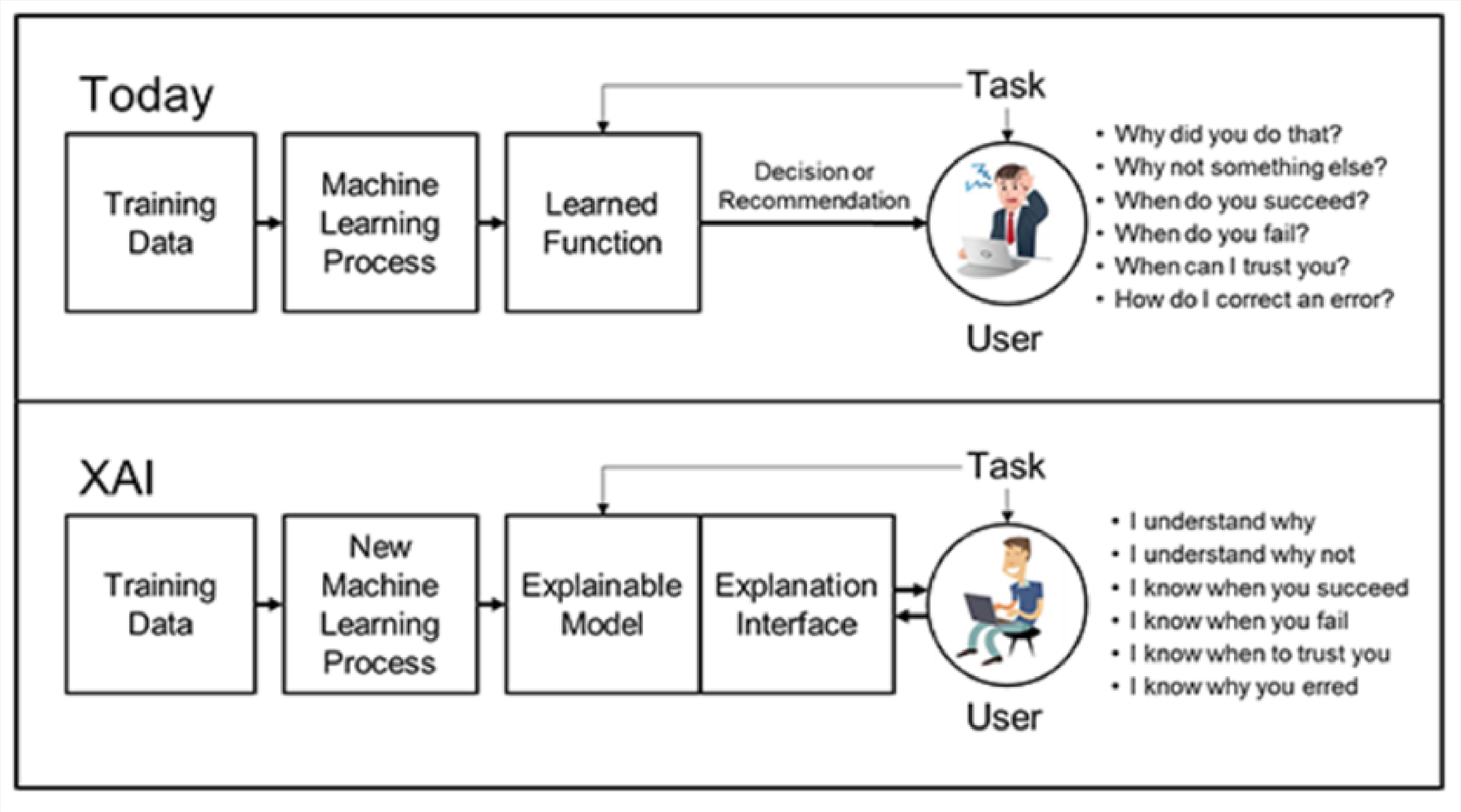

ST_Slope <= 1.5
gini = 0.493
samples = 100.0%
value = [0.441, 0.559]
class = Yes
MaxHR <= 0.036
gini = 0.288
samples = 56.9%
value = [0.175, 0.825]
class = Yes
Cholesterol <= 0.007
gini = 0.327
samples = 43.1%
value = [0.794, 0.206]
class = No
Sex <= 0.5
gini = 0.202
samples = 47.7%
value = [0.114, 0.886]
class = Yes
ChestPainType <= 0.5
gini = 0.5
samples = 9.3%
value = [0.485, 0.515]
class = Yes
FastingBS <= 0.5
gini = 0.405
samples = 5.3%
value = [0.282, 0.718]
class = Yes
ChestPainType <= 0.5
gini = 0.231
samples = 37.7%
value = [0.866, 0.134]
class = No
RestingBP <= 0.032
gini = 0.449
samples = 6.0%
value = [0.341, 0.659]
class = Yes
RestingBP <= 0.021
gini = 0.15
samples = 41.7%
value = [0.082, 0.918]
class = Yes
Cholesterol <= 0.043
gini = 0.269
samples = 3.4%
value = [0.16, 0.84]
class = Yes
RestingBP <= 0.032
gini = 0.439
samples = 5.9%
value = [0.674, 0.326]
class = No
ExerciseAngina <= 0.5
gini = 0.453
samples = 10.6%
value = [0.654, 0.346]
class = No
Oldpeak <= 2.25
gini = 0.095
samples = 27.1%
value = [0.95, 0.05]
class = No
(...)
(...)
(...)
(...)
(...)
(...)
(...)
(...)
(...)
(...)
(...)
(...)

Explainable AI

- Explainable AI (XAI) has emerged as a new paradigm in AI research that focuses on developing models that can provide a clear and understandable explanation of their decision-making process.
- The goal of XAI is to make AI systems more transparent and accountable, so that humans can trust them and make informed decisions based on their output.
- XAI aims to enhance the transparency and interpretability of AI algorithms, making them more accessible and useful for clinicians and patients.

Purpose of XAI

- Transparency
- Interpretability
- Trust
- Clinical Decison-making
- Patient outcomes



LIME

- Local Interpretable Model-Agnostic Explanations (LIME) is a popular technique in the field of Explainable AI (XAI) that can be used for analyzing and interpreting machine learning models in the context of cardiovascular disease.
- The purpose of LIME is to provide local, human-interpretable explanations of the predictions made by a black-box machine learning model.
- Local interpretability of models consists of providing detailed explanations for why an individual prediction was made.
- LIME can be used to explain the risk factors that are contributing the most to a patient's likelihood of developing a particular type of heart disease.

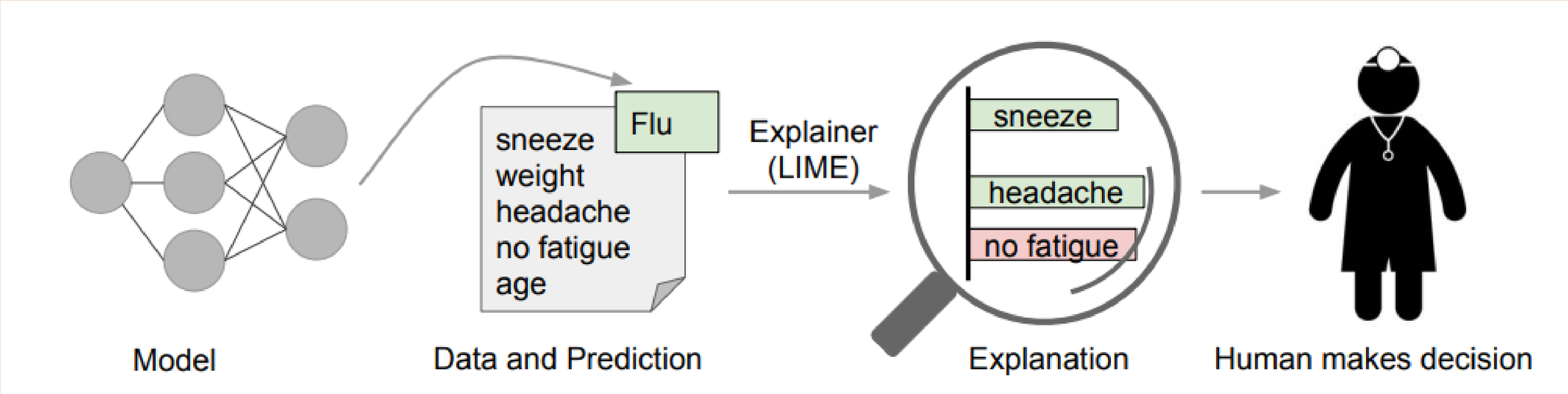
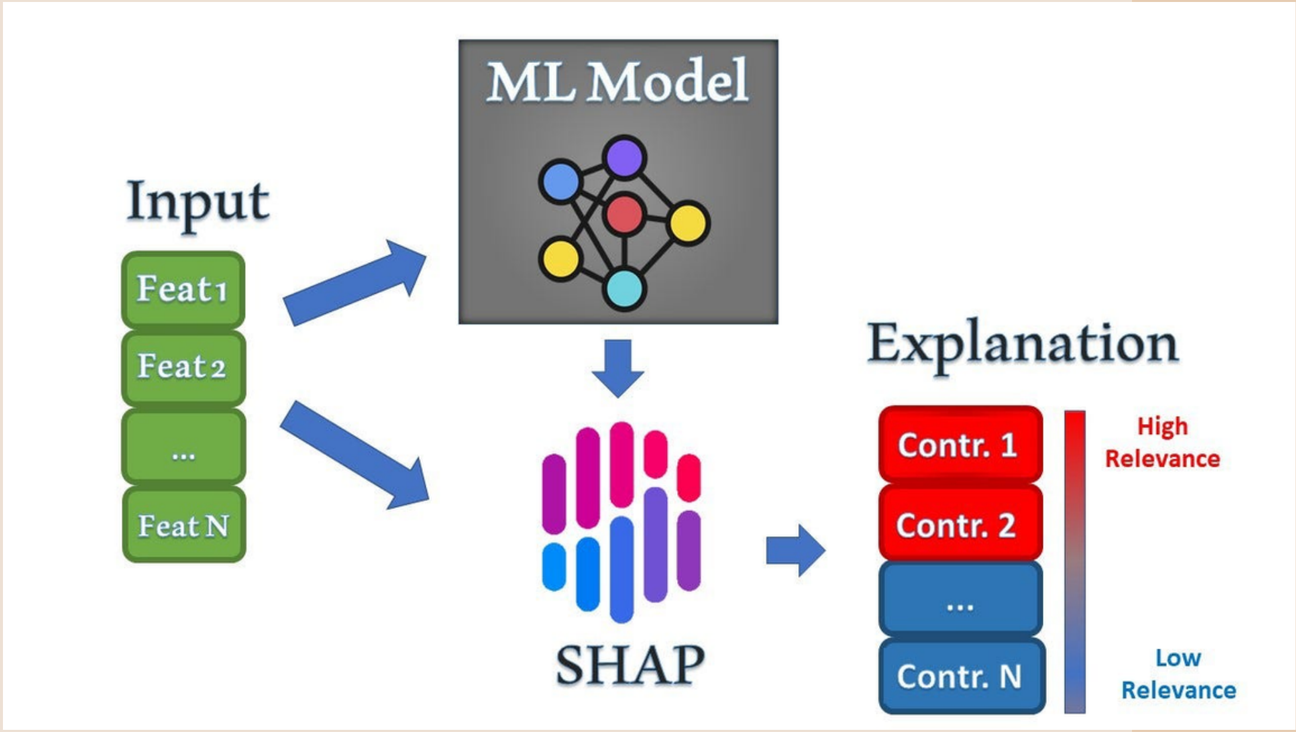


Fig. LIME

$$\xi(x) = \underset{g \in G}{\operatorname{argmin}} \quad \mathcal{L}(f, g, \pi_x) + \Omega(g)$$

SHAP

- Shapley Additive exPlanations (SHAP) values determine how each attribute contributes to the model's prediction.
- The technique assigns an importance value to each feature or risk factor based on how much it contributes to the prediction made by the model.
- In the context of cardiovascular disease analysis, SHAP can help identify the most important risk factors associated with the disease.
- By understanding the importance of these risk factors, doctors and healthcare professionals can make more informed decisions about treatment and prevention strategies for patients.



$$SHAP_{feature}(x) = \sum_{set:\, feature \in set} [|set| \times \binom{F}{|set|}]^{-1} [Predict_{set}(x) - Predict_{set \setminus feature}(x)]$$

What-if tool

- The What-If Tool (WIT) provides an easy-to-use interface for expanding understanding of a black-box classification or regression ML model.
- What-If Tool is an interactive and user-friendly visual interface used for exploring and understanding machine learning models.
- It allows users to perform "what-if" analysis on a trained model by modifying input features and observing the effect on the model's predictions in real-time.
- The What-If Tool provides several features that help users to understand how the model works and how to improve its performance. Some of the key features of the What-If Tool include:

1. Data exploration
2. Model analysis
3. Input modification
4. Counterfactual analysis

Conclusion

- We build a model which will analyse the clinical data of the individuals.
- Use Explainable AI to provide the reasoning behind predictions made by prediction model.
- Model gives us the explanations of how the ouput comes.
- It gives us the five most affecting factors that are mostly affecting the life of patient.
- It give us the visual representation which is second opinion for the doctors.
- Visual representation is more user friendly and doctors can easily explain it to the patients

Thank you