Problem Statement:

Health care: Check Heart attack possibility

The prediction of cardiac disease with use of Machine Learning (ML) and Artificial Intelligence (AI) using the provided dataset.

The Dataset

This database contains 76 attributes, but all published experiments refer to using a subset of 14 of them. In particular, the Cleveland database is the only one that has been used by ML researchers to this date. This dataset gives the information related to heart disease. Dataset contain 14 columns. It is divided into 13 predictor (independent) variables, each describing the medical details of the patient, and 1 target (dependent) variable classifying the patient into low or high chances of heart attack.

Metadata Information

Index	Column	Data Type	Description
0	AGE	Integer	Age in years
1	SEX	Binary	1 = male; 0 = female
2	CP	Categorical	Chest pain type: 0,1,2,3
3	TRESTBPS	Integer	Resting blood pressure (in mm hg)
4	CHOL	Integer	Serum cholesterol (in mg/dl)
5	FPS	Binary	Fasting blood sugar > 120 mg/dl 1 = true; 0 = false
6	RESTECH	Binary	Resting electrocardiographic results 1 = positive; 0: negative
7	THALACH	Integer	Maximum heart rate achieved
8	EXANG	Binary	Exercise induced angina 1 = yes; 0 = no
9	OLDPEAK	Float	ST depression induced by exercise relative to rest
10	SLOPE	Categorical	The slope of the peak exercise ST segment; 0,1,2
11	CA	Integer	Number of major vessels (0-3) colored by fluoroscopy
12	THAL	Categorical	0 = normal; 1 = fixed defect; 2 = reversable defect
13	TARGET	Binary	1 = high chances; 0 = low chances

Model performance

Logistic Regression

Hyperparameters used:
'Cs': 5, 'class_weight': {1: 0.4, 0: 0.6}, 'cv': 5, 'dual': False,
'fit_intercept': True, 'penalty': 'l2', 'solver': 'lbfgs'

Cross Validation Accuracy (5 Fold) = **0.89**

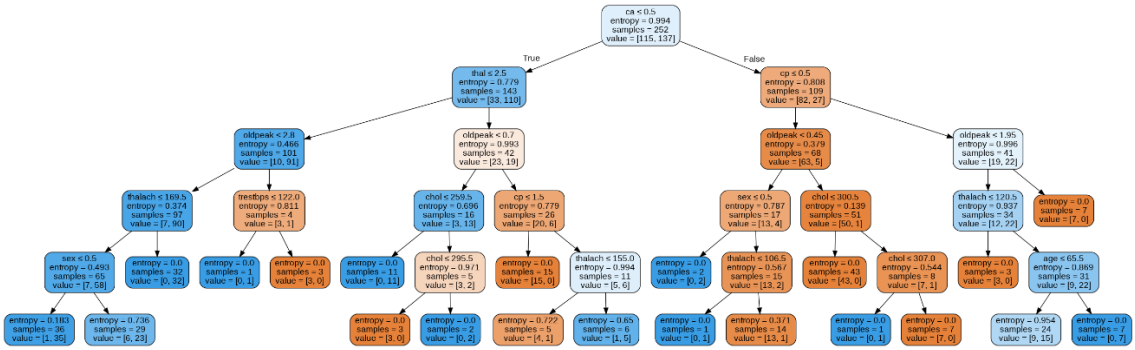
	Train	Test	Confusion Matrix for predictions			
				Predicted		
Accuracy	0.849	**0.86**	True		0	1
F1 Score	0.86	0.8679		0	20	3
Precision	0.861	0.92		1	4	23
Recall	0.8613	0.8214	Sensitivity	**0.88**	Specificity	0.8518

Decision Tree

Hyperparameters used:
criterion = 'entropy', max_depth = 5, min_samples_leaf = 3, min_samples_split = 3

	Train	Test	Confusion Matrix for predictions			
				Predicted		
Accuracy	0.9246	**0.94**	True		0	1
F1 Score	0.934	0.949		0	19	3
Precision	0.88	0.9032		1	0	28

Recall	0.985	1.0	Sensitivity	0.90	Specificity	**1**



Random Forests

Hyperparameters used:
'criterion': 'gini', 'max_depth': 75, 'max_features': 'log2', 'min_samples_leaf': 8, 'min_samples_split': 4, 'n_estimators': 150

Cross Validation Accuracy (5 Folds) = **0.85**

	Train	Test	Confusion Matrix for predictions			
				Predicted		
Accuracy	1	**0.85**	True		0	1
F1 Score	1	0.846		0	20	4
Precision	1	0.846		1	4	22
Recall	1	0.846	Sensitivity	0.846	Specificity	0.846

K Nearest Neighbours

Hyperparameters used:
algorithm= 'auto', leaf_size= 2, n_neighbors=19, weights = 'uniform'

Cross Validation Accuracy (5 Folds) = 0.79

	Train	Test	Confusion Matrix for predictions			
				Predicted		
Accuracy	0.8	**0.75**	True		0	1
F1 Score	0.83	0.78		0	19	3
Precision	0.85	0.76		1	0	28
Recall	0.825	0.755	Sensitivity	0.90	Specificity	**1**

Support Vector Machines

Hyperparameters used:
'C': 100, 'gamma': 0.0001, 'kernel': 'rbf'

Cross Validation Accuracy (5 Folds) = 0.79						
	Train	Test	Confusion Matrix for predictions			
				Predicted		
Accuracy	0.861	**0.8**	True		0	1
F1 Score	0.88	0.799		0	14	8
Precision	0.81	0.75		1	4	24
Recall	0.9635	0.85	Sensitivity	0.75	Specificity	**0.92**

Neural Networks

Cross Validation Accuracy (5 Folds) =						
	Train	Test	Confusion Matrix for predictions			
				Predicted		
Accuracy	0.7988	**0.85**	True		0	1
F1 Score	0.85	0.85		0	25	6
Precision	0.86	0.86		1	3	27
Recall	0.85	0.85	Sensitivity		Specificity	