Harnessing Supervised Machine Learning for Cardiovascular Disease Analysis and Prediction

Applied Data Science – Final Presentation

Professor : Dr Amir Hesam Salavati

Presenter : Ali Zahedzadeh

mail: alizahedzadeh7@gmail.com

TeIAS | Tehran Institute for Advanced Studies

- Cardiovascular Disease

- Supervised Machine learning

(Random Forest, Decision Tree, Logistic Regression and etc.)

Research Article

Supervised Machine Learning-Based Cardiovascular Disease Analysis and Prediction

M. D. Amzad Hossen,[1] **Tahia Tazin**,[1] **Sumiaya Khan**,[1] **Evan Alam**,[1] **Hossain Ahmed Sojib**,[1] **Mohammad Monirujjaman Khan**,[1] **and Abdulmajeed Alsufyani**[2]

[1]*Department of Electrical and Computer Engineering, North South University, Bashundhara, Dhaka 1229, Bangladesh*
[2]*Department of Computer Science, College of Computers and Information Technology, Taif University, P.O. Box 11099, Taif 21944, Saudi Arabia*

Correspondence should be addressed to Mohammad Monirujjaman Khan; monirujjaman.khan@northsouth.edu

Received 10 November 2021; Revised 23 November 2021; Accepted 26 November 2021; Published 10 December 2021

Academic Editor: Ewa Rak

About Dataset

Heart disease dataset

Column	Description
Age	Age of the patient in completed years
Sex	Gender of the patient
Cp	Chest Pain type (1: typical angina, 2: atypical angina, 3: non-anginal pain, 4: asymptomatic)
Trestbps	Resting blood pressure (in mm Hg)
Chol	Cholesterol in mg/dl fetched via BMI sensor
FBS	Fasting blood sugar > 120 mg/dl (1 = true; 0 = false)
Resting	resting electrocardiographic results
Thalach	Maximum heart rate achieved
Exang	Exercise-induced angina (1 = yes, 0 = no)
Oldpeak	Previous peak
Ca	Number of major vessels (0-4)
Thal	0 = Normal, 1 = Fixed, 2 = Reversible, 3 = Non-Reversible
Target	0 = Less chance of heart attack, 1 = More chance of heart attack

UCI ML Repository
https://archive.ics.uci.edu

About Dataset

- No null Values

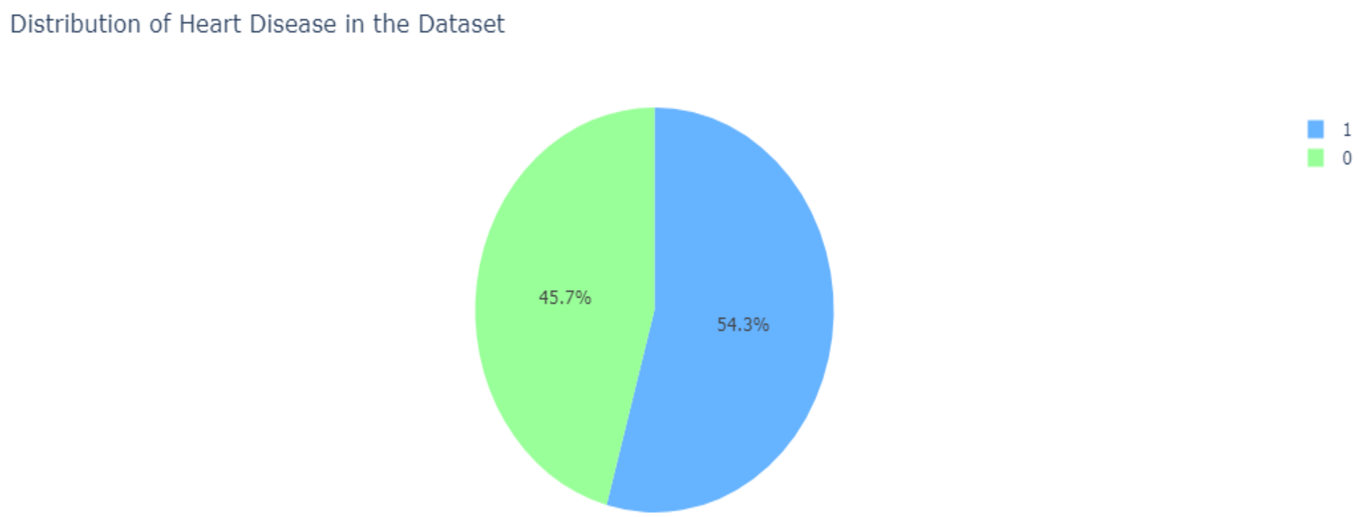
```
heart_data.isnull().sum()
```

[38]

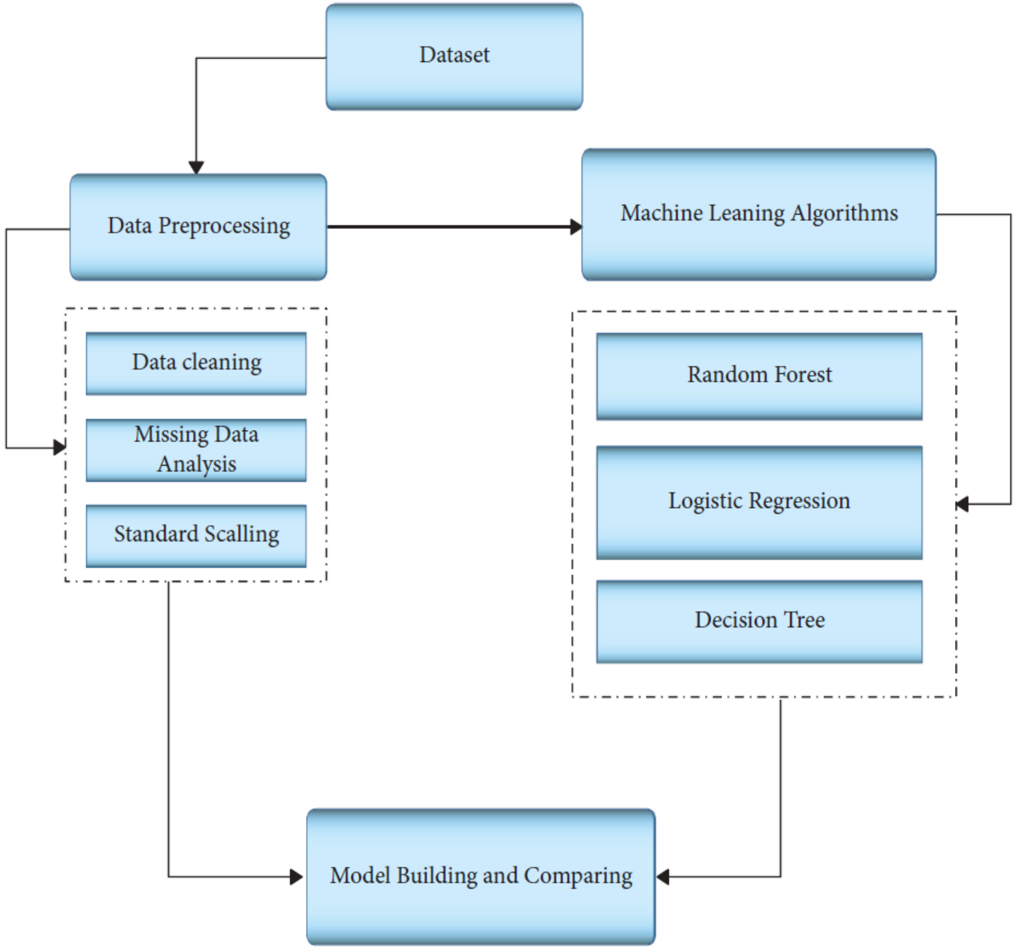
```
Age         0
Sex         0
Cp          0
Trestbps    0
Chol        0
FBS         0
Resting     0
Thalach     0
Exang       0
Oldpeak     0
Slope       0
Ca          0
Thal        0
Target      0
dtype: int64
```

- Balanced Dataset

165 cardiac disease & **138** noncardiac disease



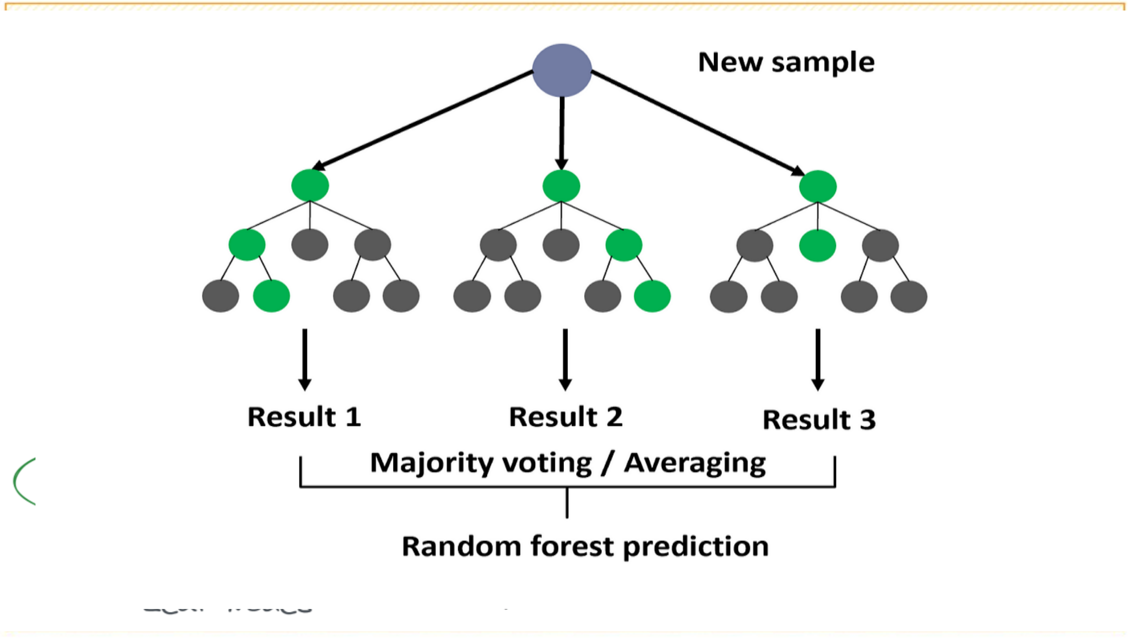
Methodology



System’s schematic diagram

Methodology

- Logistic Regression
- Decision Tree
- Random Forest



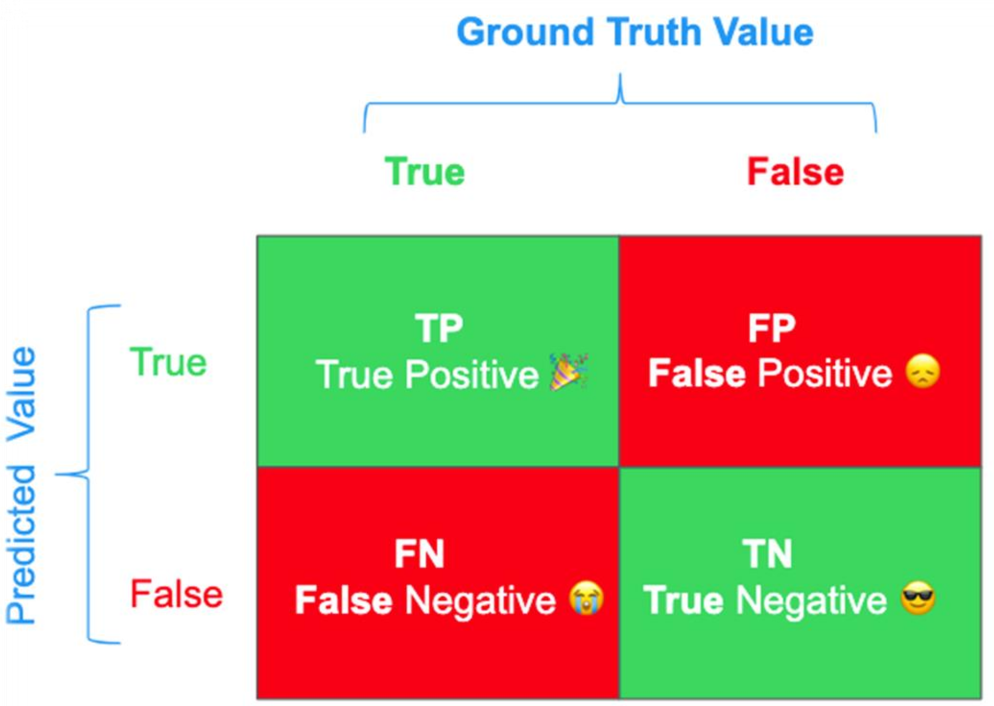
About Metrics

$$\boldsymbol{Precision} = \frac{TP}{TP + FP}$$

$$\boldsymbol{Recall} = \frac{TP}{TP + FN}$$

$$\boldsymbol{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}$$

$$\boldsymbol{F1 - Score} = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

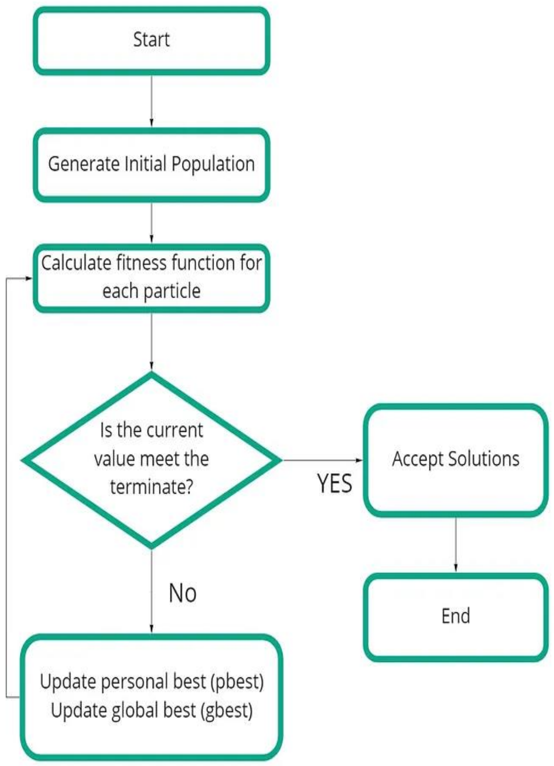


Related Works

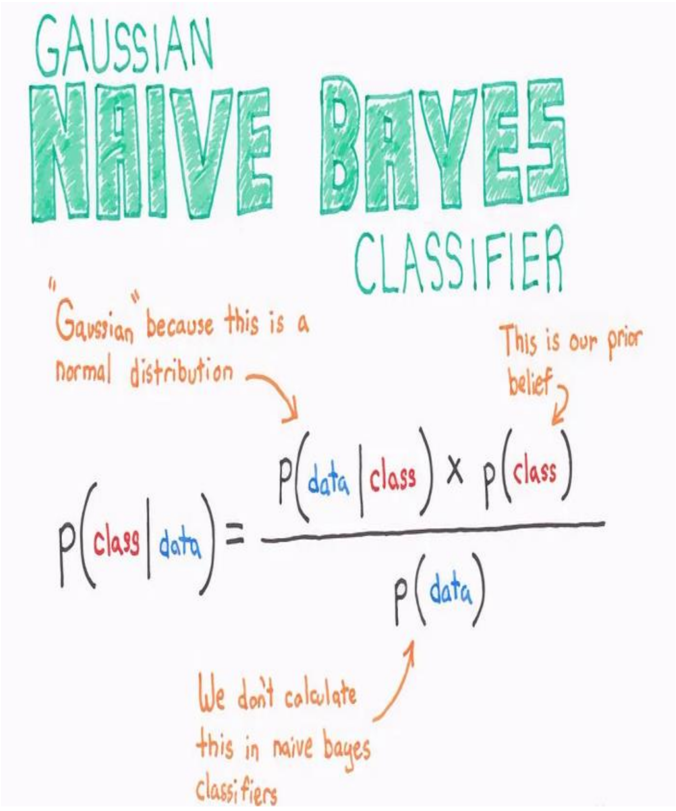
Prediction system for heart disease using Naive Bayes and particle swarm optimization
January 2018

- Particle Swarm Optimization (PSO)
- Naïve Bayes Classifier

PSO algorithm



Naïve Bayes Classifier

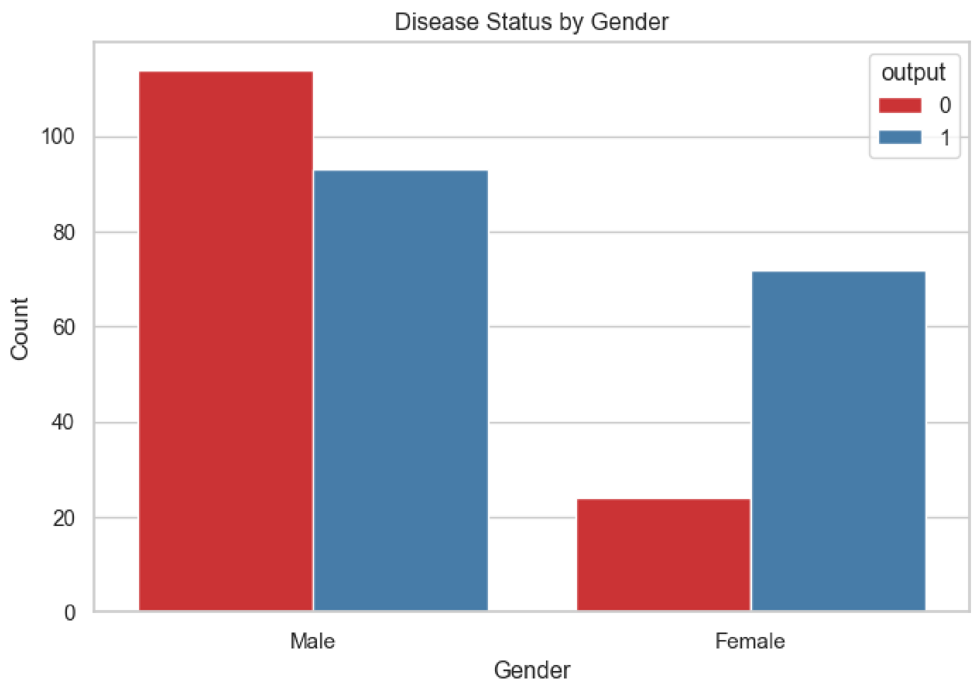


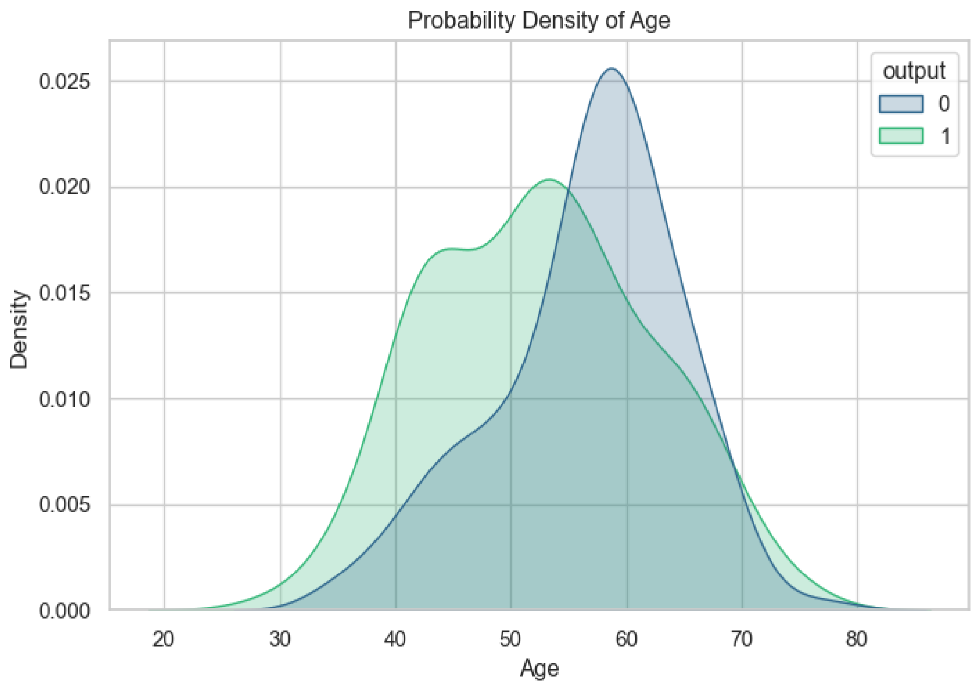
Paper Result

TABLE 2: Classification result of the three different models.

Model	Precision (%)	Recall (%)	*F*1-score (%)	Accuracy (%)
Random forest	77	87	82	80
Decision tree	71	74	72	72
Logistic regression	92	92	92	92

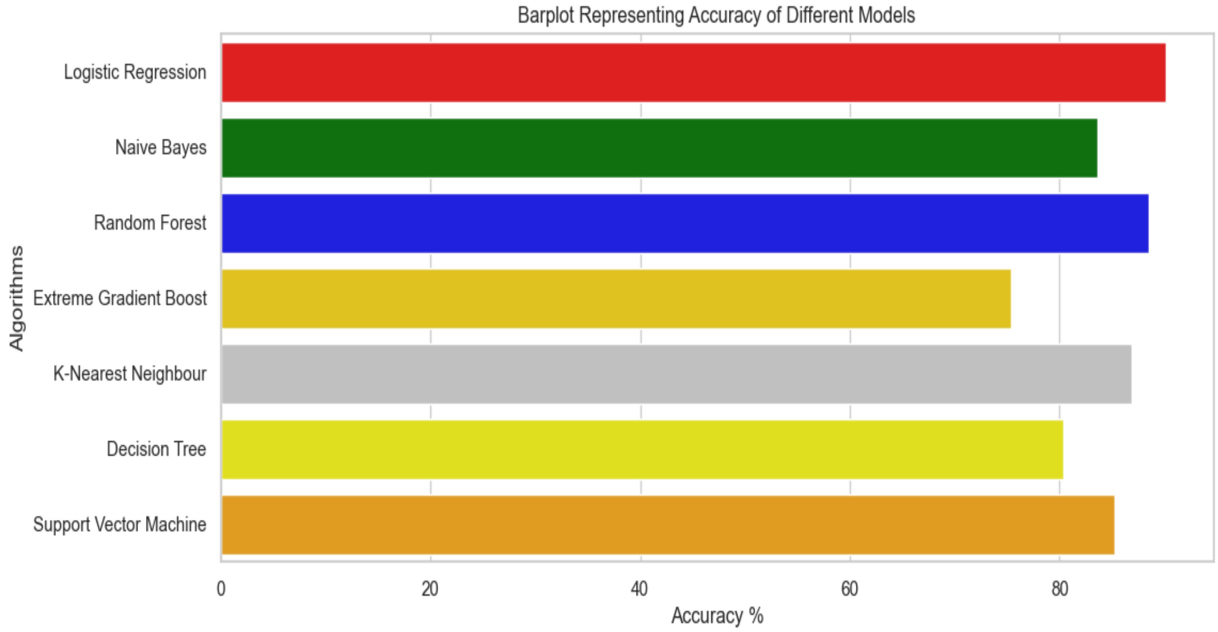
Reproduced Result - Data Analysis





Reproduced Result - Models Evaluation

Model	Accuracy
Logistic Regression	90.163934
Naive Bayes	83.606557
Random Forest	88.524590
Extreme Gradient Boost	75.409836
K-Nearest Neighbour	86.885246
Decision Tree	80.327869
Support Vector Machine	85.245902



- The study emphasizes the potential impact of the developed diagnostic system using machine learning in predicting heart disease, with the possibility of early identification of at-risk patients and improved accuracy in diagnosing cardiac abnormalities.

- Logistic Regression is identified as the best model for predicting heart disease due to its consistently higher accuracy, precision, and overall performance compared to other machine learning algorithms (**Based on Paper Result**)

- Rapid and cost-effective heart disease prediction using accessible dataset and evolving machine learning algorithms, offering the potential to significantly impact public health by identifying at-risk individuals and contributing to a reduction in the rising death rate.

- Other models can be used for this study that have acceptable accuracy (**Based on Reproduce Result**)

mail: alizahedzadeh7@gmail.com

Refrences

https://www.kaggle.com/datasets/redwankarimsony/heart-disease-data

https://archive.ics.uci.edu

https://www.popai.pro

https://www.geeksforgeeks.org/understanding-logistic-regression

https://towardsdatascience.com/3-techniques-to-avoid-overfitting-of-decision-trees-1e7d3d985a09

https://www.researchgate.net/

https://medium.com/@roiyeho/random-forests-98892261dc49

Thank you for your attention