JERUSALEM COLLEGE OF ENGINEERING,CHENNAI-100

DEPARTMENT OF INFORMATION TECHNOLOGY

HEART SIGHT USING ML

PRESENTED BY

- SOWMIYA (1307212046)
- SUVALAKSHMI (130721205048)

SUPERVISED BY

DR.K.SUNDARAMOORTHY

HOD IT

OBJECTIVE

- Heart attack prediction is one of the serious causes of morbidity in the world's population and the clinical data analysis includes a very crucial disease.

- By adding some advance algorithm in Machine Learning we create a high risk of prediction in Heart Attack.

ABSTRACT

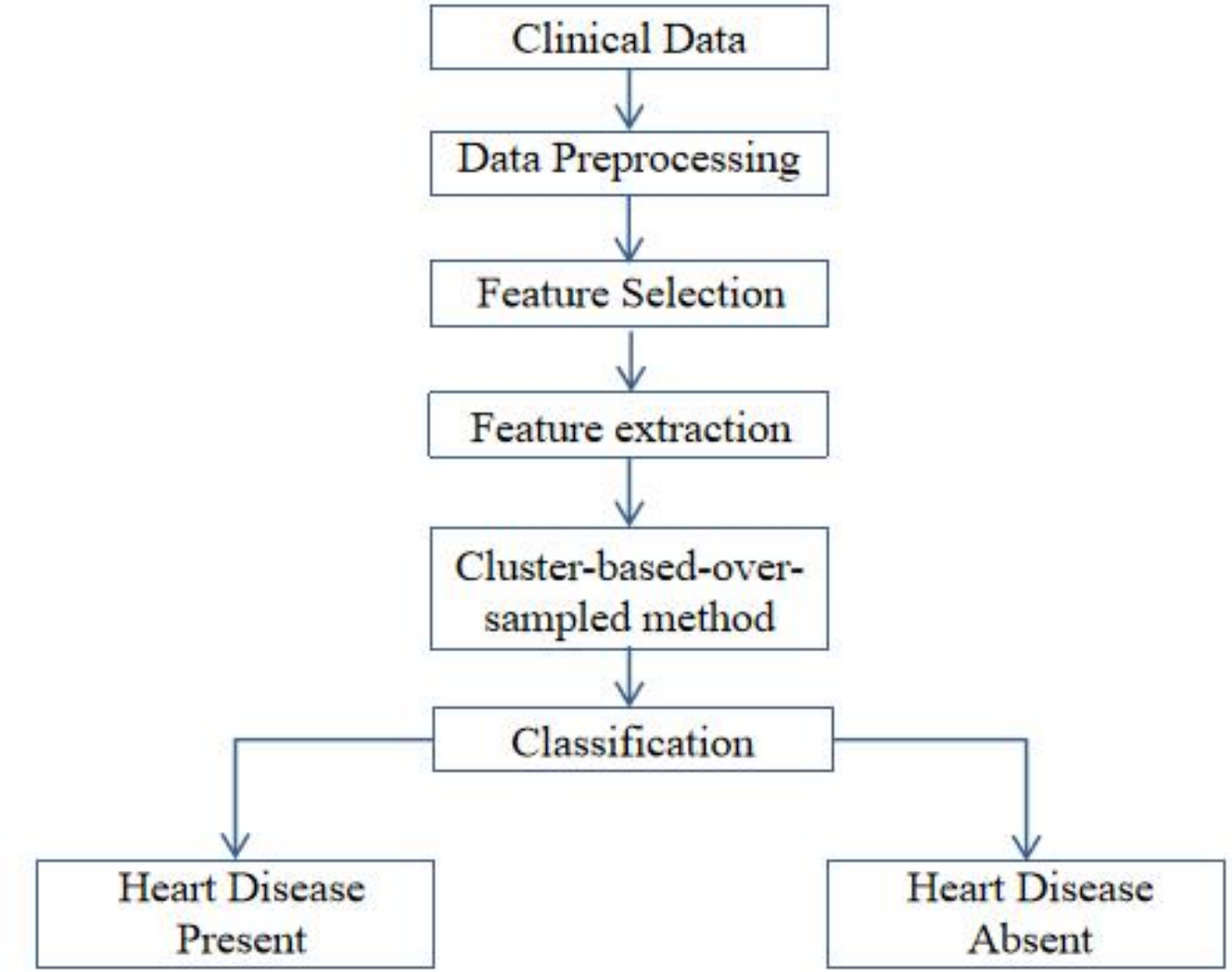
- Designing an effective system for Heart Attack Risk Prediction and this will be more accuracy to find the person in danger of heart attack at any time.

- The dataset will catch the datas of the people using pre-processed datasets available in medical database.

- This dataset will show age, sex, cholestrol, blood pressure, blood sugar, heart rate of the person to address the risk of heart attack.

- Using a ML algorithm the output will predict the high rate of heart attack risk by graph points.

ARCHITECHTURE DIAGRAM



LITERATURE SURVEY

- C.Boukhatem, H.Y.Youssef and A.B.Nassif, "Heart Disease Prediction Using Machine Learning,"2022 Advances in Science and Engineering Technology International Conferences(ASET), Dubai, United Arab Emirates, 2022, pp.1-6, doi:10.1109/ASET53988.2022.9734880.

  LINK:https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9734880&isnumber=9734306

  TOPIC- Development of Heart Attack Prediction using SVM algorithm

- V. Sharma, S.Yadav and M.Gupta, "Heart Disease Prediction using Machine Learning Techniques,"2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), Greater Noida, India, 2020, pp.177-181, doi:10.1109/ICACCCN51052.2020.9362842.

  LINK:https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9734880&isnumber=9734306

  TOPIC- Presdiction of Heart Attack using Decision Tree in ML.

EXISTING SYSTEM

- The Existing system represents the Heart Attack risk of the people with low prediction using such algorithms.

- Normally the existing system of the prediction will present using Support Vector Machine(SVM), Decision Tree, Naive Bayes algorithms.

- This algorithms may predict the risk of people with low percentages this will lead people to confuse and stress themselves about their health.

- The final graph will show the people risk zone using given algorithms at low prediction.

PROPOSED SYSTEM

•In this system we adding the boosting algorithms in machine learning to get enhanced and high prediction of the risk of heart disease.

•Ada-Boost Classifier, Hist Gradient Boost Classifier and using some algorithms we compare with the datasets and gives the result with high predictions and this will predict the people risk zone perfectly.

•We predict comparing such algorithms and present the high frequency prediction percentage by graph and this may help the person to alert precautionary about their health.

•Hence the risk of the people will get low soon.

PHASE - I
MODULE-1 PRE-PROCESSING DATA

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	age	sex	cp	trtbps	chol	fbs	restecg	thalachh	exng	oldpeak	slp	caa	thall	output
2	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
3	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
4	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
5	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
6	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1
7	57	1	0	140	192	0	1	148	0	0.4	1	0	1	1
8	56	0	1	140	294	0	0	153	0	1.3	1	0	2	1
9	44	1	1	120	263	0	1	173	0	0	2	0	3	1
10	52	1	2	172	199	1	1	162	0	0.5	2	0	3	1
11	57	1	2	150	168	0	1	174	0	1.6	2	0	2	1
12	54	1	0	140	239	0	1	160	0	1.2	2	0	2	1
13	48	0	2	130	275	0	1	139	0	0.2	2	0	2	1
14	49	1	1	130	266	0	1	171	0	0.6	2	0	2	1
15	64	1	3	110	211	0	0	144	1	1.8	1	0	2	1
16	58	0	3	150	283	1	0	162	0	1	2	0	2	1
17	50	0	2	120	219	0	1	158	0	1.6	1	0	2	1
18	58	0	2	120	340	0	1	172	0	0	2	0	2	1
19	66	0	3	150	226	0	1	114	0	2.6	0	0	2	1
20	43	1	0	150	247	0	1	171	0	1.5	2	0	2	1
21	69	0	3	140	239	0	1	151	0	1.8	2	2	2	1
22	59	1	0	135	234	0	1	161	0	0.5	1	0	3	1
23	44	1	2	130	233	0	1	179	1	0.4	2	0	2	1
24	42	1	0	140	226	0	1	178	0	0	2	0	2	1
25	61	1	2	150	243	1	1	137	1	1	1	0	2	1
26	40	1	3	140	199	0	1	178	1	1.4	2	0	3	1
27	71	0	1	160	302	0	1	162	0	0.4	2	2	2	1
28	59	1	2	150	212	1	1	157	0	1.6	2	0	2	1
29	51	1	2	110	175	0	1	123	0	0.6	2	0	2	1
30	65	0	2	140	417	1	0	157	0	0.8	2	1	2	1
31	53	1	2	130	197	1	0	152	0	1.2	0	0	2	1
32	41	0	1	105	198	0	1	168	0	0	2	1	2	1
33	65	1	0	120	177	0	1	140	0	0.4	2	0	3	1
34	44	1	1	130	219	0	0	188	0	0	2	0	2	1
35	54	1	2	125	273	0	0	152	0	0.5	0	1	2	1
36	51	1	3	125	213	0	0	125	1	1.4	2	1	2	1
37	46	0	2	142	177	0	0	160	1	1.4	0	0	2	1
38	54	0	2	135	304	1	1	170	0	0	2	0	2	1
39	54	1	2	150	232	0	0	165	0	1.6	2	0	3	1
40	65	0	2	155	269	0	1	148	0	0.8	2	0	2	1

- Data preprocessing is the process of transforming raw data into an understandable form.
- Preprocessing of data is mainly to check the data quality like Accuracy, Completeness, Consistency, Timeliness, Believability, Interpretability.
- CSV is a file format which allows us to save the tabular data, such as spreadsheets. It is useful for huge datasets and can use these datasets in programs.

PARAMETER AGE & GENDER

GENDER	LOW RISK	HIGH RISK
MALE	<45	>45
FEMALE	<45	>55

GENDER CRITERIA

MALE :

Men have higher levels of testosterone, but levels decrease with age. Low testosterone levels are linked to heart disease.

FEMALE :

Women have higher levels of estrogen hormone before mensuration, which help keep blood vessels healthy. so after the end period of mensuration the risk will get high.

BLOOD PRESSURE LEVEL

BLOOD PRESSURE CATEGORY	SYSTOLIC mm Hg (upper number)		DIASTOLIC mm Hg (lower number)
NORMAL	<120	and	>80
ELEVATED	120-129	and	>80
HIGH BLOOD PRESSURE (HYPERTENSION) STAGE 1	130-139	or	80-89
HIGH BLOOD PRESSURE (HYPERTENSION) STAGE 2	140 or HIGHER	or	90 or HIGHER
HYPERTENSIVE CRISIS (consult the doctor)	HIGHER THAN 180	and/or	HIGHER THAN 120

CHOLESTROL RANGE

CHOLESTROL LEVEL	TOTAL CHOLESTROL	LDL CHOLESTROL	HDL CHOLESTROL
HEALTHY	UNDER 200	UNDER 100	60 AND HIGHER
AT RISK	200-239	100-159	M:49-59 F:50-59
DANGEROUS	240 AND HIGHER	160 AND HIGHER	M:UNDER 40 F:UNDER 50

THALASSEMIA RANGE

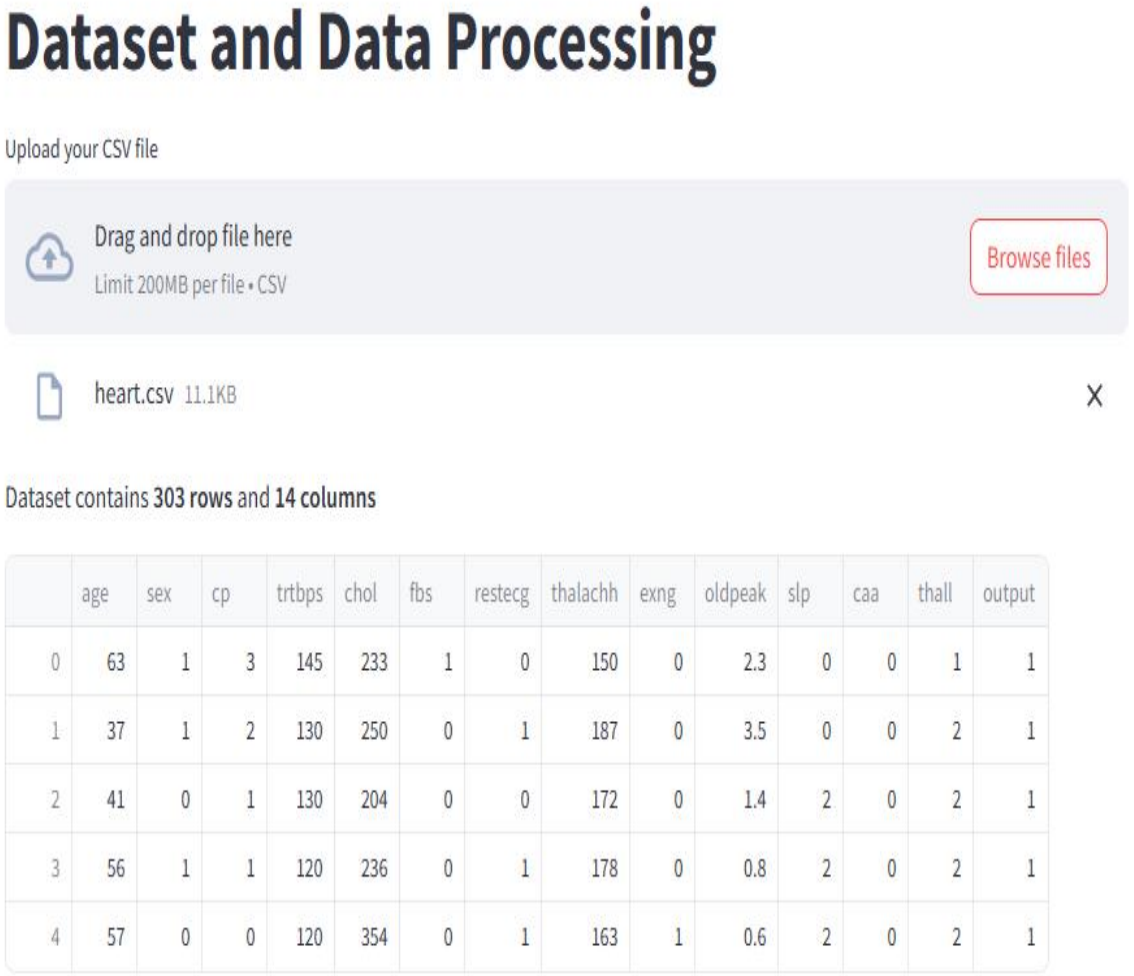
NORMAL HB VALUE:

MALE: 13.5 gms - 17 gms

FEMALE: 12.0 gms - 15.5 gms

THALASSEMIA LEVELS	HB LEVEL	MCV (mean corpuscular volume)	MCH (mean corpuscular hemoglobin)
BETA MINOR	M:11-15 g/dl F:10-13 g/dl	-	-
INTERMEDIA	7-10 g/dl	50 and 80 fl	16-24 pg
MAJOR	LESS THAN 7	50 and 70 fl	12 and 20 pg

OUTPUT OF PRE-PROCESSING DATA



	age	sex	cp	trtbps	chol	fbs	restecg	thalachh	exng	oldpeak	slp	caa	thall	output
0	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
1	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
2	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
3	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
4	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1

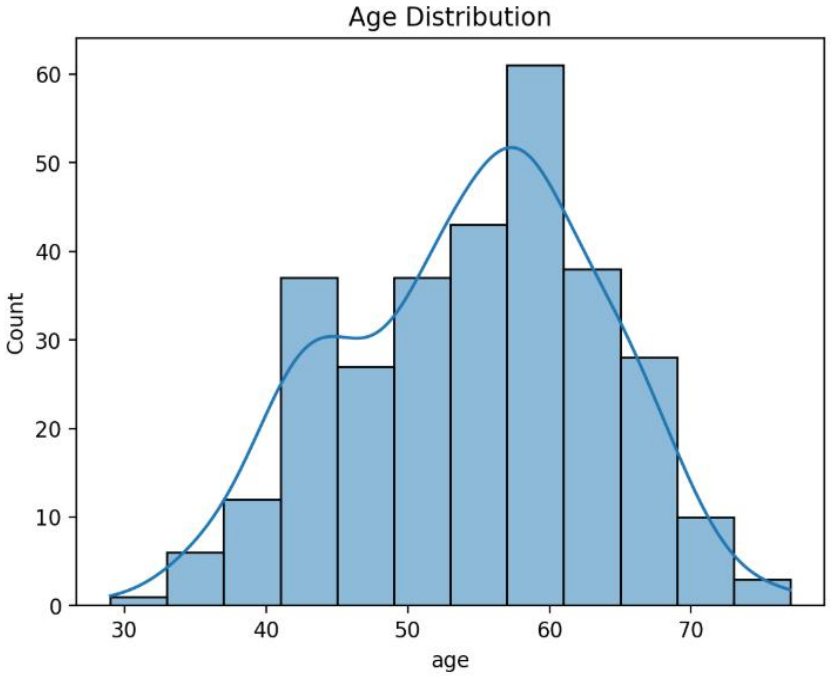
- Datasets for heart disease have been collected using Kaggle platform and store it as CSV file.
- Using Streamlit framework the dataset will processed by extracting collected data.
- Here we got the output and verification for the collected datasets.

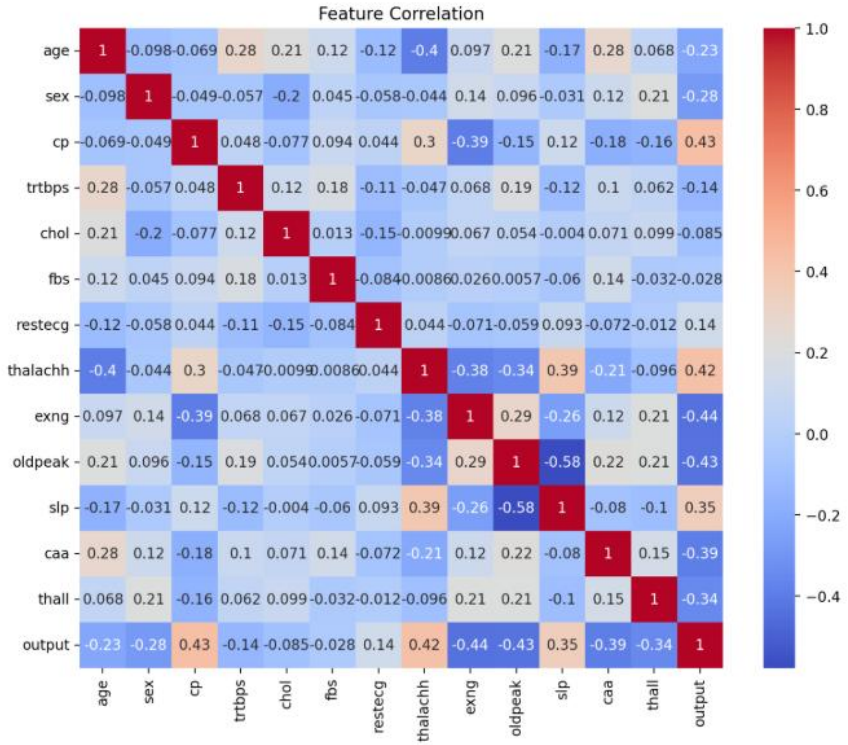
MODULE-2 FEATURE SELECTION

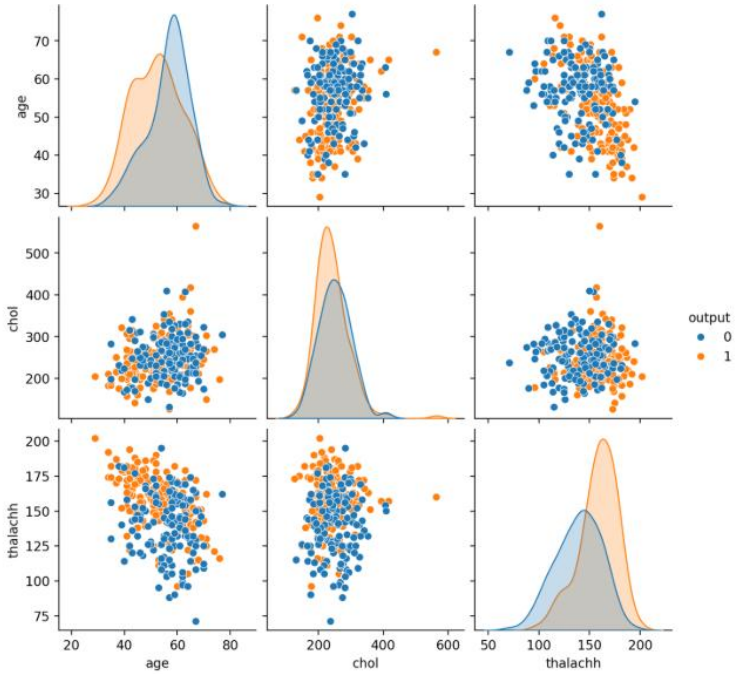
Feature name	Features
var 1	age
var 2	sex
var 3	chest pain type (4 values)
var 4	testing blood pressure
var 5	serum cholesterol in mg/dL
var 6	fasting blood sugar > 120 mg/dL
var 7	resting electrocardiographic results (values 0,1,2)
var 8	maximum heart rate achieved
var 9	excersice induced angina
var 10	old peak = ST depression induced

- Feature Selection is nothing but selecting the features for required high accuracy.
- By Pre-processing datas we get age, sex, BP(Blood Pressure) level, Diabeties level, Heart rate, Cholestrol level and etc.
- Using a requied features will helps to achieve a correct accuracy level in heart disease prediction.

OUTPUT OF SELECTED FEATURES







PHASE - II
MODULE 1 MODEL TRAINING

- Ada Boost Classifier:

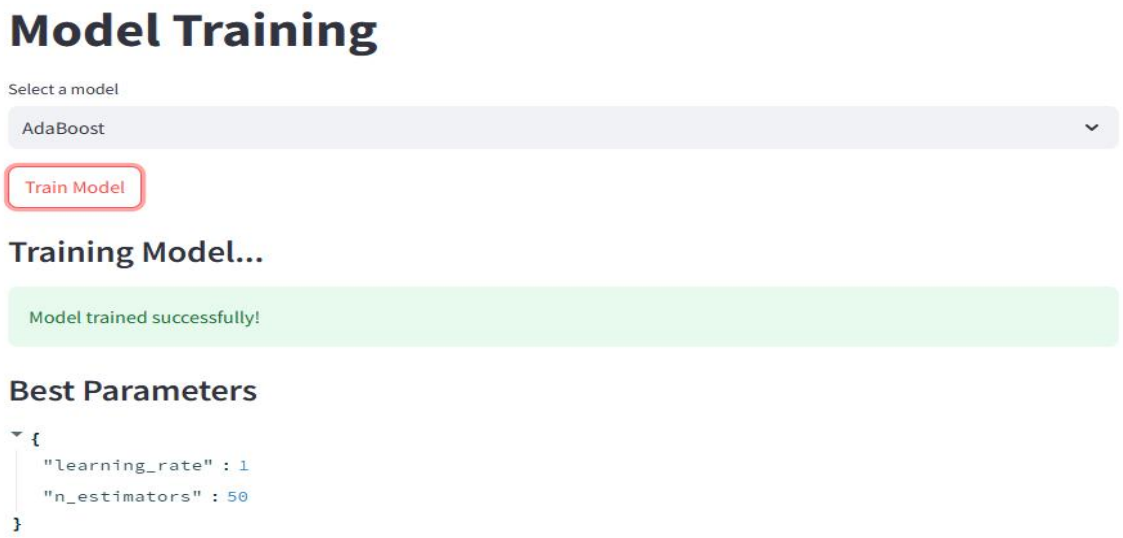
Train a weak classifier (e.g., decision tree) on the weighted samples. It improve the accuracy of prediction by transforming weak learners, typically (one level decision trees) into a collective strong classifier.

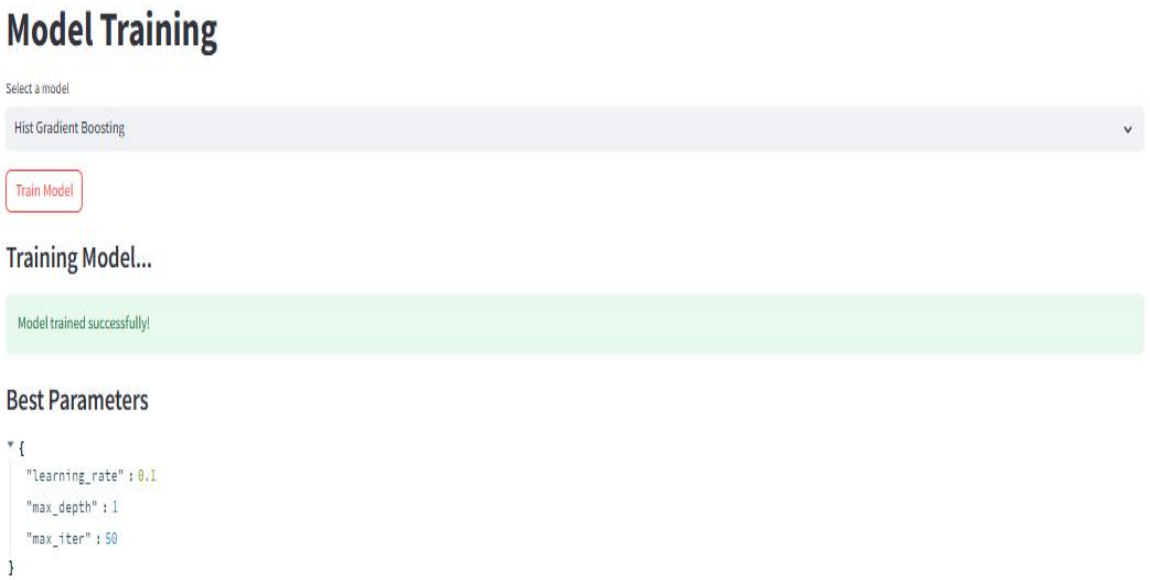
- Hist Gradient Boosting Classifier:

It improve the efficiency of the traditional gradient boosting algorithm by discretizing continuous input features into bins (histograms).

- These two classifier can be used for both classification and regression tasks due to its high predictive accuracy.

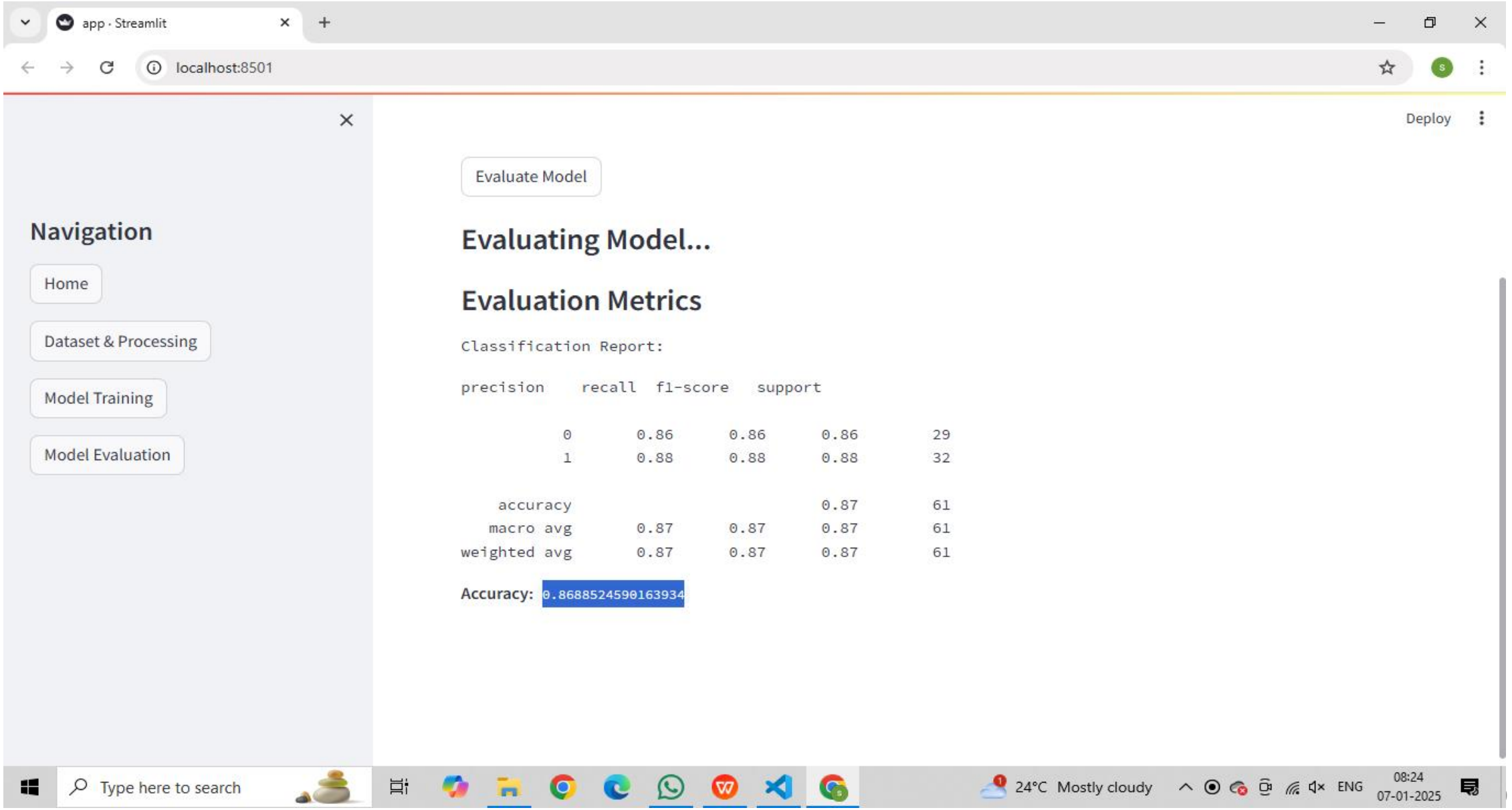
OUTPUT OF MODEL TRAINING

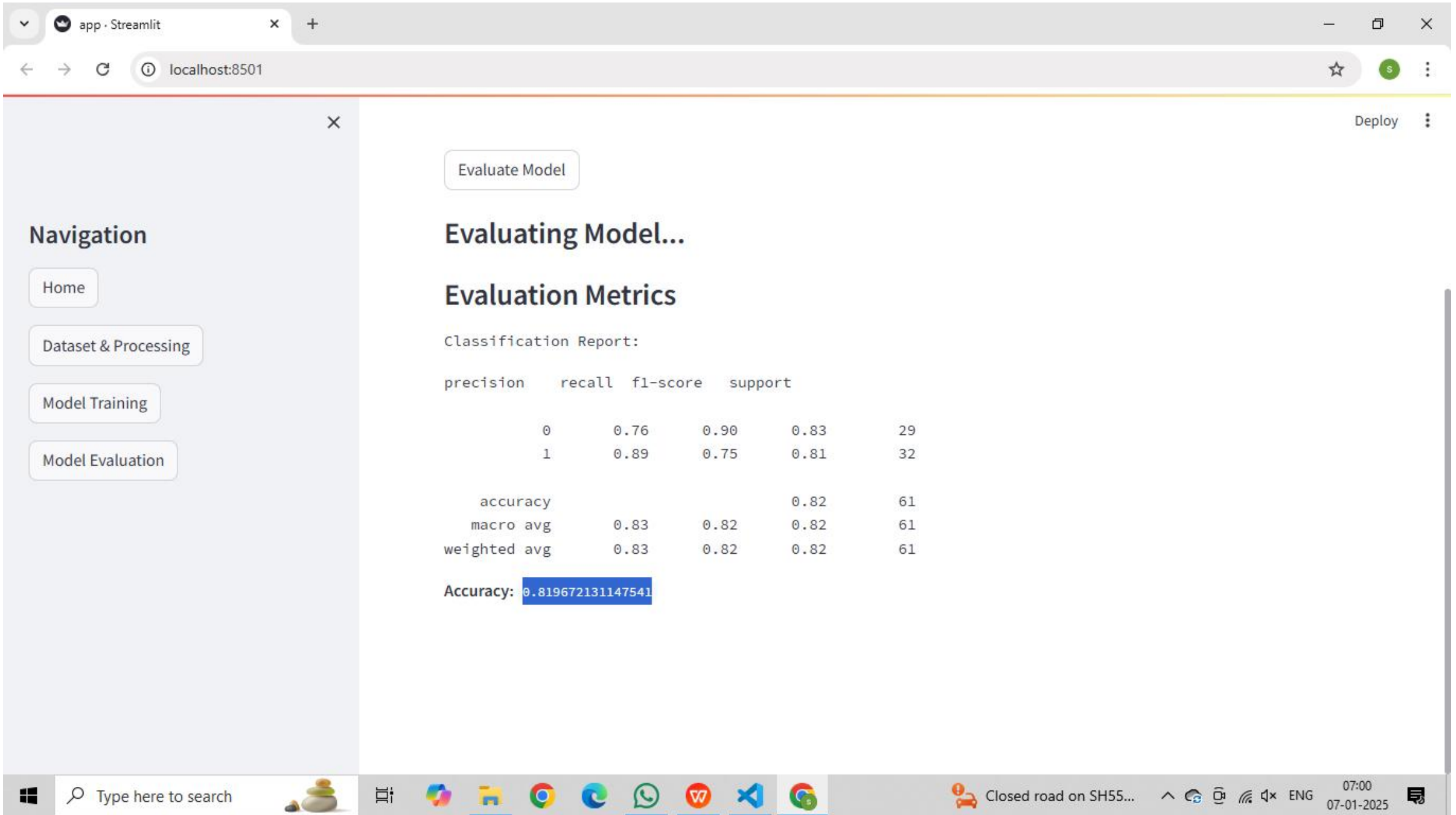




- This picture shows the implementation of different algorithms comparison of prediction risk of heart disease.
- By comparing these we got an high risk factor predicting algorithms which is Hist Gradient Classifier and Ada Boost Classifier.

FINAL OUTPUT




app · Streamlit
localhost:8501
Deploy
Navigation
Home
Dataset & Processing
Model Training
Model Evaluation
Evaluate Model
Evaluating Model...
Evaluation Metrics
Classification Report:
precision recall f1-score support
0 0.76 0.90 0.83 29
1 0.89 0.75 0.81 32
accuracy 0.82 61
macro avg 0.83 0.82 0.82 61
weighted avg 0.83 0.82 0.82 61
Accuracy: 0.819672131147541
Type here to search
Closed road on SH55...
ENG
07:00
07-01-2025

THANK YOU !