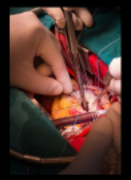
CARDIOVASCULAR RISK ASSESSMENT

ETE | JULY 2024
GUIDE: DR. DINESH SHARMA
SCHOOL OF INFORMATION AND TECHNOLOGY

Problem Statement

Overview of Heart Disease



Understanding Heart Disease

Heart Disease refers to the various conditions that affect the heart's functioning and structure.

It is one of the leading causes of mortality worldwide and needs to be addressed immediately.

Understanding Heart Disease involves:

- recognizing its types
- factors
- symptoms
- diagnostic methods
- treatment options
- preventive measures.

CONCLUSION
&
FUTURE SCOPE

Machine Learning Techniques

Importance of Data in Predictive Modeling

WEB APPLICATION

Case Study

Exploring a Real-World Application of Machine Learning in Predicting Heart Disease

CARDIOVASCULAR RISK ASSESSMENT

ETE | JULY 2024

GUIDE: DR. DINESH SHARMA
SCHOOL OF INFORMATION AND TECHNOLOGY

Understanding Heart Disease

Heart Disease refers to the various conditions that affect the heart's functioning and structure.

It is one of the leading causes of mortality worldwide and needs to be addressed immediately.

Understanding Heart Disease involves:

- recognizing its types
- factors
- symptoms
- diagnostic methods
- treament options
- preventive measures.

Overview of Heart Disease

Heart disease refers to a range of conditions that affect the heart's functioning including:

- Coronary Artery Disease
- Heart Attacks
- Heart Failure
- Arrhytmias

Problem Statement

- The impact of Heart Disease is multifaceted, affecting individuals, families, healthcare system, and economies.

- Statistics suggest that in 2019 globally an estimated 17.9 million people died from cardiovascular diseases.

Definition of Machine Learning

Machine learning is the subset of Artificial Intelligence that involves developement of models and algorithms that enables computers to perform tasks like prediction, classification, analysis and many more without explicit instructions

Machine Learning Techniques

Machine learning Techniques can be broadly classified into three major categories:

- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning

Importance of ML in Healthcare

- Machine Learning is transforming healthcare in numerous ways by enhancing the accuracy of diagnosis, personalizing treatments, and improving patient outcomes.

- Few Key Areas include:
  - Diagnosis
  - Personalised Medicine
  - Medical Imaging
  - Automated Health Recording
  - Clinical Analysis

```
[3]: {'data': {'ids': None,
  'features':       age  sex  cp  trestbps  chol  fbs  restecg  thalach  exang  oldpeak  \
  0      63    1   1       145   233    1        2      150      0      2.3
  1      67    1   4       160   286    0        2      108      1      1.5
  2      67    1   4       120   229    0        2      129      1      2.6
  3      37    1   3       130   250    0        0      187      0      3.5
  4      41    0   2       130   204    0        2      172      0      1.4
  ..    ...  ...  ..       ...   ...  ...      ...      ...    ...      ...
  298    45    1   1       110   264    0        0      132      0      1.2
  299    68    1   4       144   193    1        0      141      0      3.4
  300    57    1   4       130   131    0        0      115      1      1.2
  301    57    0   2       130   236    0        2      174      0      0.0
  302    38    1   3       138   175    0        0      173      0      0.0

       slope   ca  thal
  0        3  0.0   6.0
  1        2  3.0   3.0
  2        2  2.0   7.0
  3        3  0.0   3.0
  4        1  0.0   3.0
  ..     ...  ...   ...
  298      2  0.0   7.0
  299      2  2.0   7.0
  300      2  1.0   7.0
  301      2  1.0   3.0
  302      1  NaN   3.0

  [303 rows x 13 columns],
  'targets':      num
  0      0
  1      2
  2      1
  3      0
  4      0
  ..   ...
  298    1
  299    2
  300    3
  301    1
  302    0

  [303 rows x 1 columns],
```

Importance of Data in Predictive Modeling

High-quality data is the foundation of accurate predictive modeling in healthcare. It determines the effectiveness and reliability of machine learning algorithms in predicting heart disease.

The UC Irvene dataset for Heart Disease is a well regarded dataset in the field of Machine Learning and Cardiovascular Research

It is widely used as a benchmark for developing and testing new algorithms in machine learning and data mining.

Feature Extraction for ML Modeling

A crucial step in preparing data for machine learning modeling that helps enhance the accuracy and interpretability of Machine Learning models.

Involves Steps like:

- Understanding the Dataset
- Data cleaning
- Data Preprocessing
- Feature Engineering

Analysing The Target Variable

```
# Count the occurrences of each unique value in the target variable
target_count = data.target.value_counts()

# Print the count of each unique value in the target variable
print(target_count)

colors = ['orange','green']
target_count.plot(kind='bar',color=colors)
plt.xlabel('Target')
plt.ylabel('Count')
plt.title('Distribution of Target Variable')
plt.xticks([0, 1], ['Target 0', 'Target 1'], rotation=0)
plt.show()
```

```
target
1    165
0    138
Name: count, dtype: int64
```

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	
count	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303
mean	54.366337	0.683168	0.966997	131.623762	246.264026	0.148515	0.528053	149.646865	0.326733	1.039604	1.399340	0.729373	2.313531	0
std	9.082101	0.466011	1.032052	17.538143	51.830751	0.356198	0.525860	22.905161	0.469794	1.161075	0.616226	1.022606	0.612277	0
min	29.000000	0.000000	0.000000	94.000000	126.000000	0.000000	0.000000	71.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0
25%	47.500000	0.000000	0.000000	120.000000	211.000000	0.000000	0.000000	133.500000	0.000000	0.000000	1.000000	0.000000	2.000000	0
50%	55.000000	1.000000	1.000000	130.000000	240.000000	0.000000	1.000000	153.000000	0.000000	0.800000	1.000000	0.000000	2.000000	1
75%	61.000000	1.000000	2.000000	140.000000	274.500000	0.000000	1.000000	166.000000	1.000000	1.600000	2.000000	1.000000	3.000000	1
max	77.000000	1.000000	3.000000	200.000000	564.000000	1.000000	2.000000	202.000000	1.000000	6.200000	2.000000	4.000000	3.000000	1

Important Features and Train and Test Split

For enhanced prediction and analysis identifying important features and setting up proper train-test split are crucial steps.

For better model training and testing we generally split the data in train test split in the ration of either **"80 : 20"** or **"70 : 30"**.

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
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

---> Splitting the dataset into Training and Test split using sklearn

```
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(X, y, test_size = 0.2, random_state = 1)
```

ML Models

--> Logistic Regression
--> KNN
--> SVM
--> Naive Bayes Classifier
--> Decision Tree
--> Random Forest
--> XG Boost

DL Model

--> Neural Networks

Machine Learning Models for Heart Disease Prediction

Choosing the right model involves considering factors like the size and quality of your dataset, interpretability of results, computational resources available, and the specific requirements of our application (e.g., real-time prediction or model explainability).

The project utilizes a total of 7 Machine Learning models and 1 Deep Learning model.

ML Models

--> Logistic Regression

--> KNN

--> SVM

--> Naive Bayes Classifier

--> Decision Tree

--> Random Forest

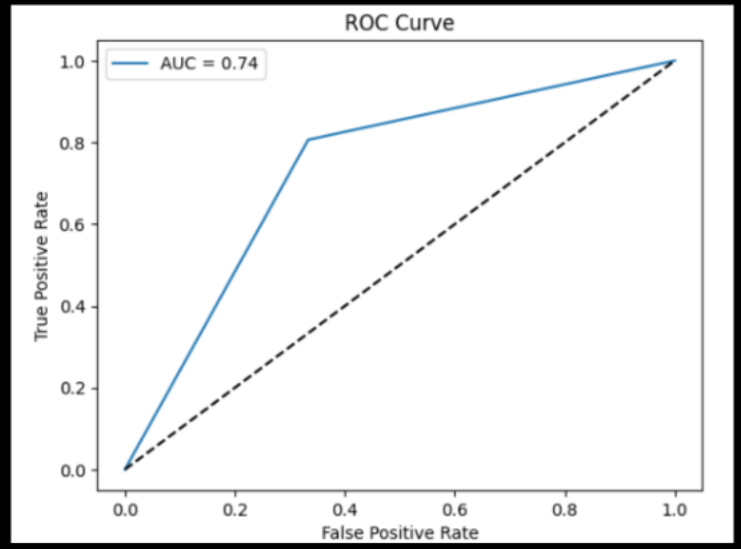
--> XG Boost

DL Model

--> Neural Networks

Logistic Regression

	precision	recall	f1-score	support
0	0.77	0.67	0.71	30
1	0.71	0.81	0.76	31
accuracy			0.74	61
macro avg	0.74	0.74	0.74	61
weighted avg	0.74	0.74	0.74	61



Logistic regression is a statistical model that analyzes the relationship between a dependent variable, like heart disease, and one or more independent variables like age, blood pressure, and cholesterol levels.

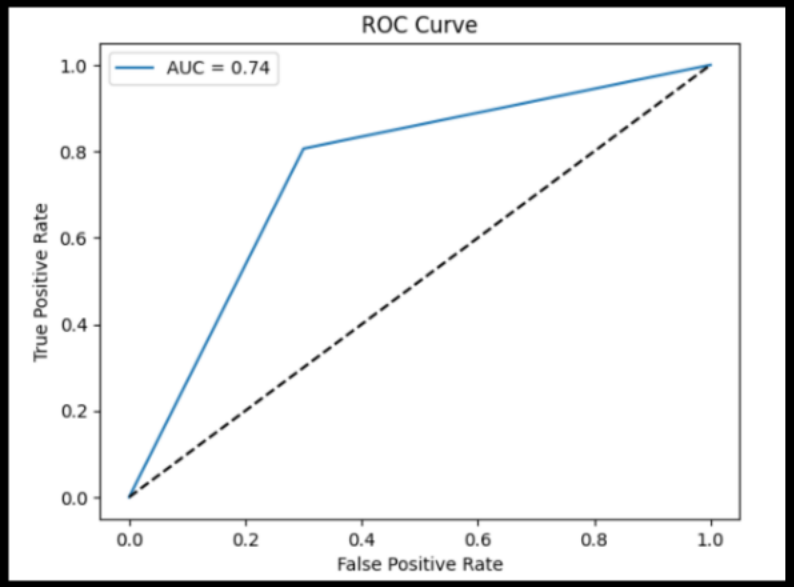
Suitable for binary classification tasks, such as predicting presence or absence of Heart Disease

K Nearest Neighbors

K-Nearest Neighbors (KNN) is a simple yet effective algorithm for heart disease prediction. It classifies new instances based on the majority class of their nearest neighbors in the training set.

KNN's non-parametric nature makes it flexible for datasets with complex relationships.

	precision	recall	f1-score	support
0	0.78	0.70	0.74	30
1	0.74	0.81	0.77	31
accuracy			0.75	61
macro avg	0.76	0.75	0.75	61
weighted avg	0.76	0.75	0.75	61

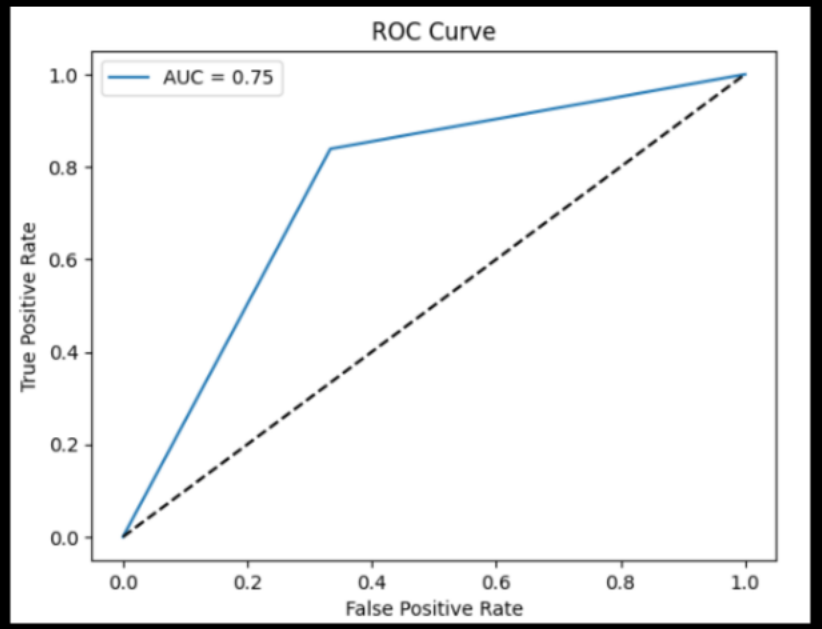


Support Vector Machines

Support Vector Machines are supervised learning models that analyze data for classification and regression tasks, making them valuable for predicting heart disease based on patterns in patient data.

Seperates datapoints into different classes,thus is effective for both linear and nonlinear classification tasks.

	precision	recall	f1-score	support
0	0.80	0.67	0.73	30
1	0.72	0.84	0.78	31
accuracy			0.75	61
macro avg	0.76	0.75	0.75	61
weighted avg	0.76	0.75	0.75	61

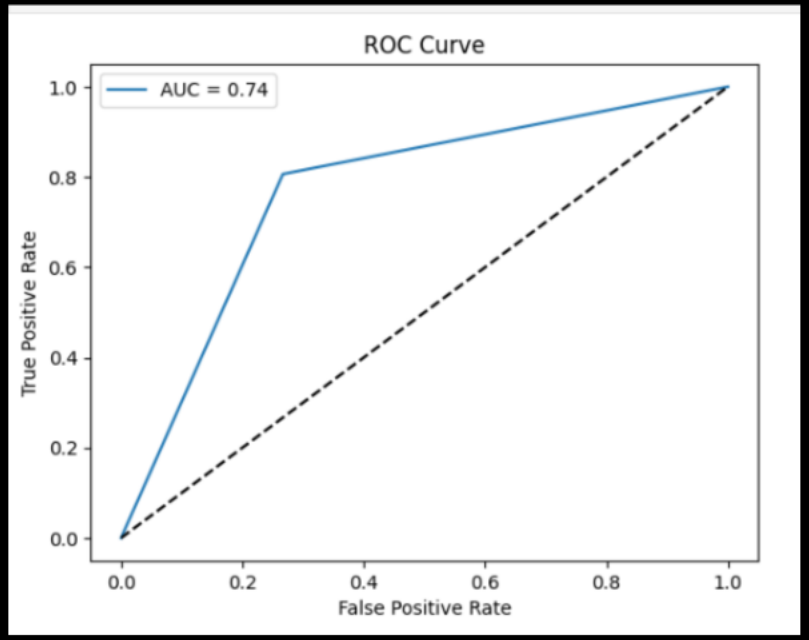


Naive Bayes Classifier

The Naive Bayes classifier is used in heart disease prediction by calculating the probability of disease presence or absence based on independent features such as age, sex, cholesterol levels, and ECG results.

Naive Bayes is computationally efficient, works well with high-dimensional data, and requires relatively few training data.

	precision	recall	f1-score	support
0	0.79	0.73	0.76	30
1	0.76	0.81	0.78	31
accuracy			0.77	61
macro avg	0.77	0.77	0.77	61
weighted avg	0.77	0.77	0.77	61

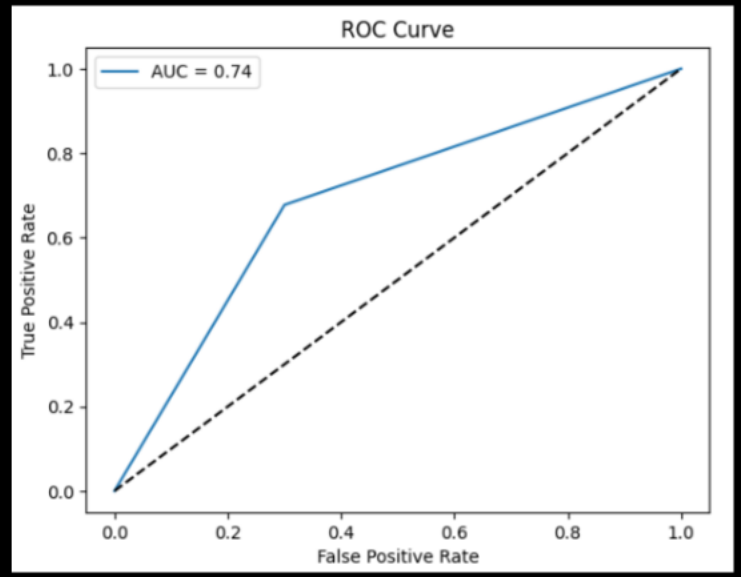


Decision Trees

Decision Trees are used for heart disease prediction by splitting data based on feature thresholds to classify instances. They can handle non-linear relationships and interactions between features effectively.

Decision Trees are interpretable, making it easy to understand how predictions are made.

	precision	recall	f1-score	support
0	0.68	0.70	0.69	30
1	0.70	0.68	0.69	31
accuracy			0.69	61
macro avg	0.69	0.69	0.69	61
weighted avg	0.69	0.69	0.69	61

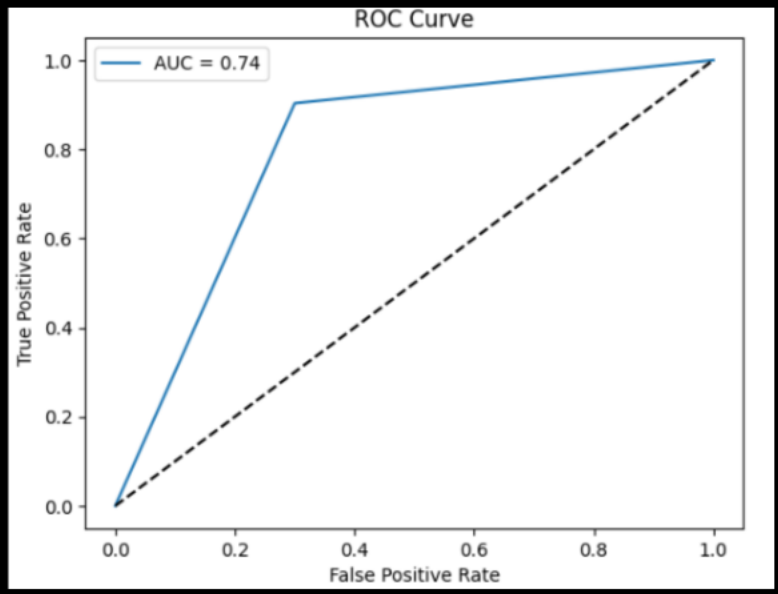


Random Forest

Random Forest, an ensemble learning method, harnesses the power of decision trees to amplify the accuracy of heart disease predictions.

Random Forest is effective in heart disease prediction by aggregating predictions from multiple decision trees trained on random subsets of data and features. It handles non-linear relationships and interactions well, making it robust for complex datasets like those in medical diagnostics.

	precision	recall	f1-score	support
0	0.88	0.70	0.78	30
1	0.76	0.90	0.82	31
accuracy			0.80	61
macro avg	0.82	0.80	0.80	61
weighted avg	0.81	0.80	0.80	61



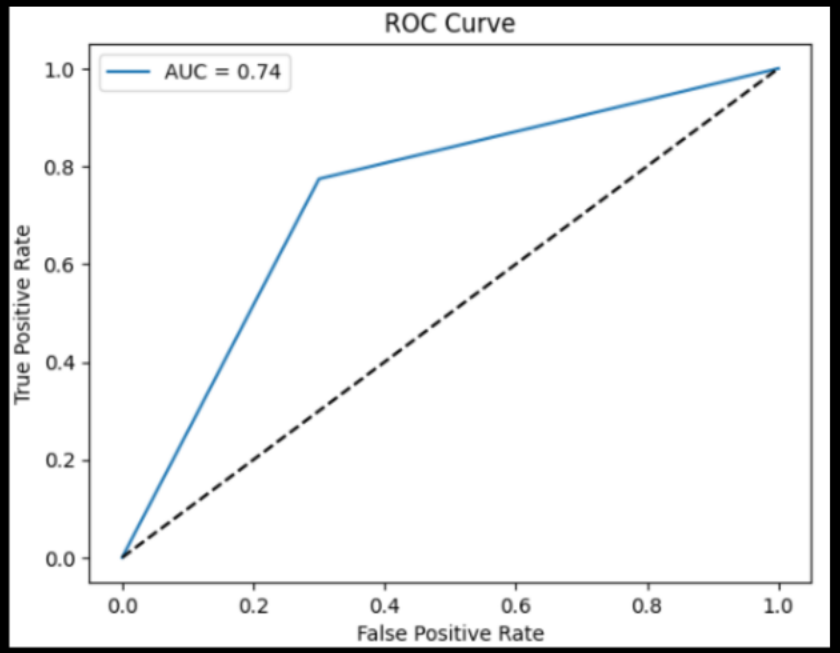
XG Boost

(Extreme Gradient Boosting) is powerful for heart disease prediction, enhancing decision tree-based models by boosting ensemble learning with gradient boosting techniques.

XGBoost handles missing data, feature interactions, and non-linearity effectively, offering robust performance.

It optimizes prediction accuracy by sequentially adding models that correct errors of previous ones.

	precision	recall	f1-score	support
0	0.75	0.70	0.72	30
1	0.73	0.77	0.75	31
accuracy			0.74	61
macro avg	0.74	0.74	0.74	61
weighted avg	0.74	0.74	0.74	61



Neural Networks

Neural Networks are applied to heart disease prediction by learning complex patterns from data through layers of interconnected nodes (neurons).

Neural Networks excel in capturing intricate relationships among features, potentially outperforming traditional models in predictive accuracy.

```
from sklearn.metrics import accuracy_score

score_nn = round(accuracy_score(y_pred_nn, y_test) * 100, 2)

print("The accuracy score achieved using Neural Network is: " + str(score_nn) + " %")
```

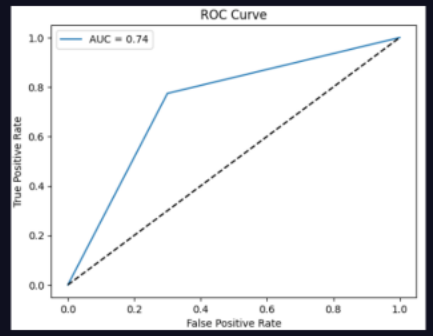
```
The accuracy score achieved using Neural Network is: 72.13 %
```

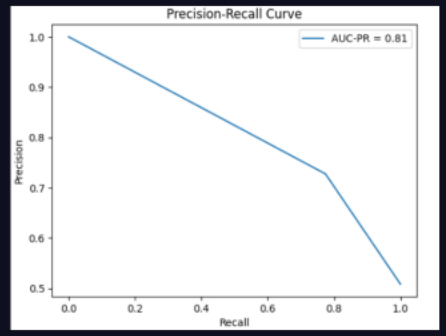
Evaluate Model Performance

Accurately assessing the performance of machine learning models is crucial in predicting heart disease.Performance evaluation in machine learning assesses how well a model predicts outcomes compared to actual data.

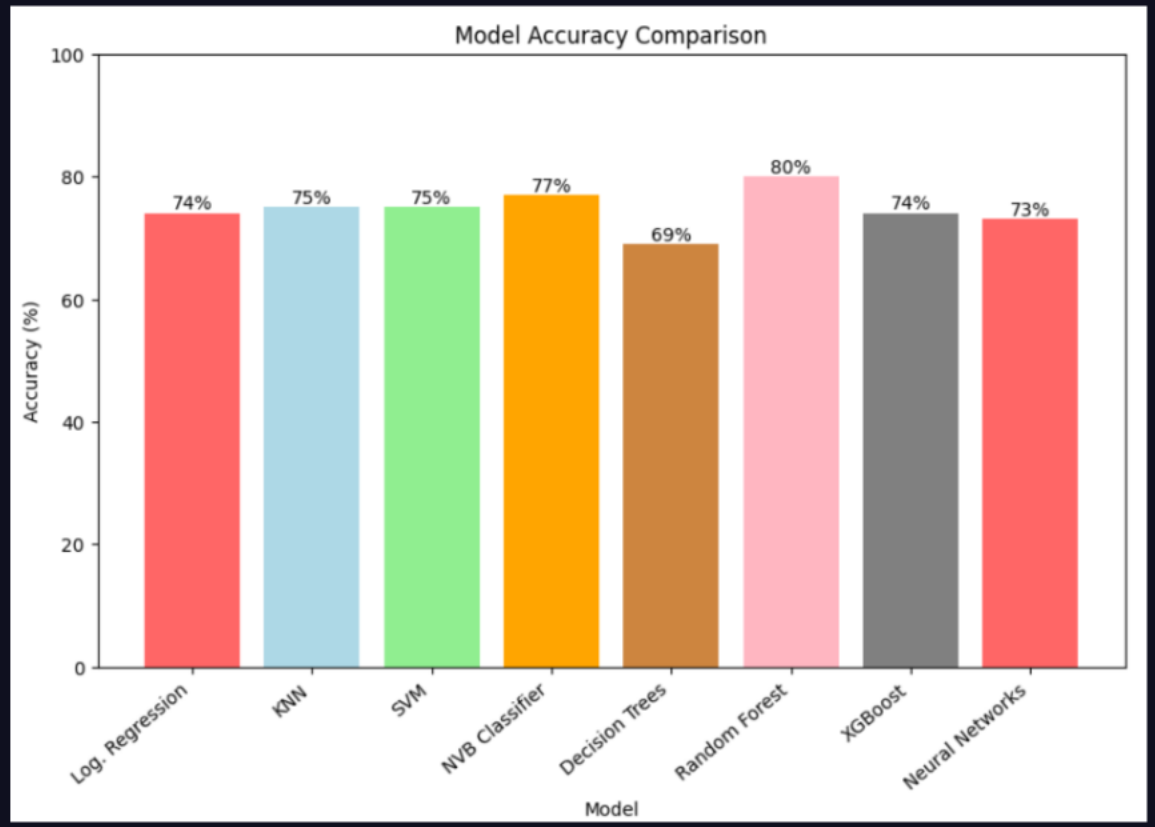
Metrics like accuracy measure overall correctness, precision gauges exactness of positive predictions, recall assesses the model's ability to find all positives, and F1 score balances precision and recall. The confusion matrix displays correct and incorrect predictions by class, providing a detailed breakdown of model performance.

	precision	recall	f1-score	support
0	0.77	0.67	0.71	30
1	0.71	0.81	0.76	31
accuracy			0.74	61
macro avg	0.74	0.74	0.74	61
weighted avg	0.74	0.74	0.74	61



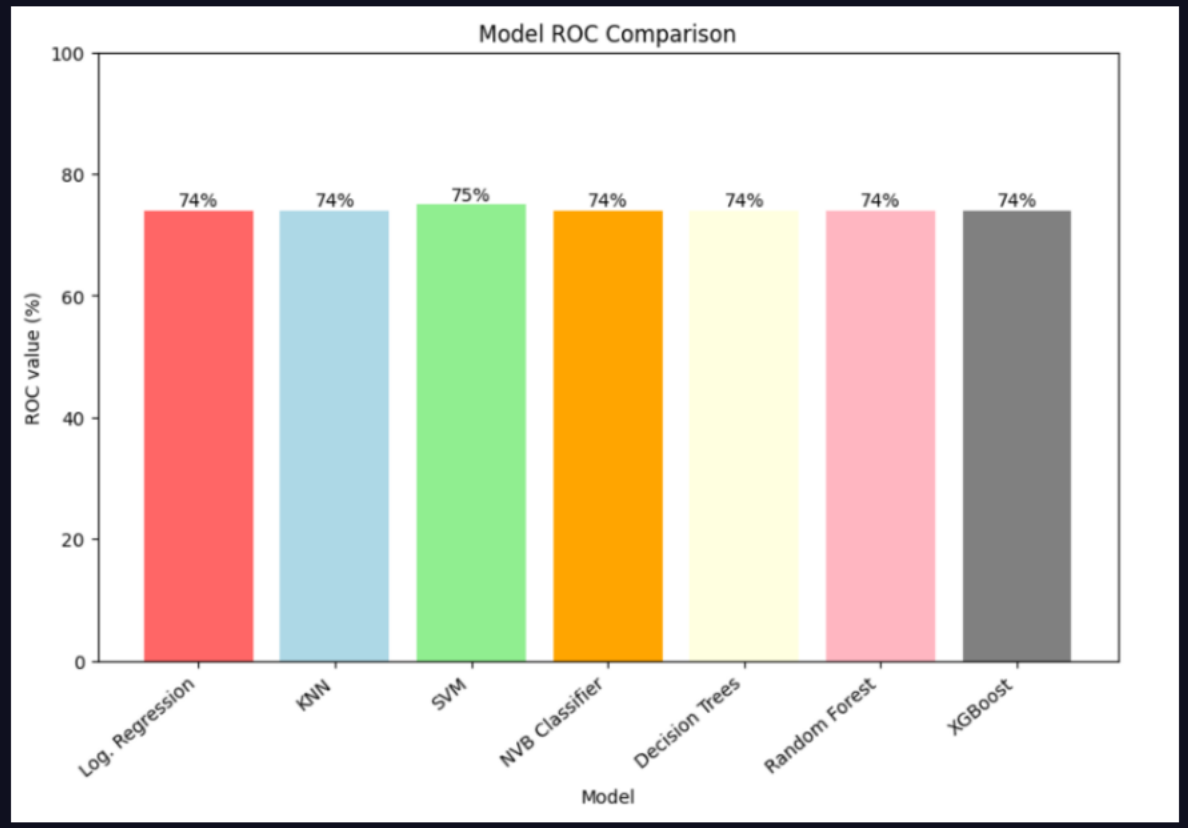


Accuracy, Precision, and Recall

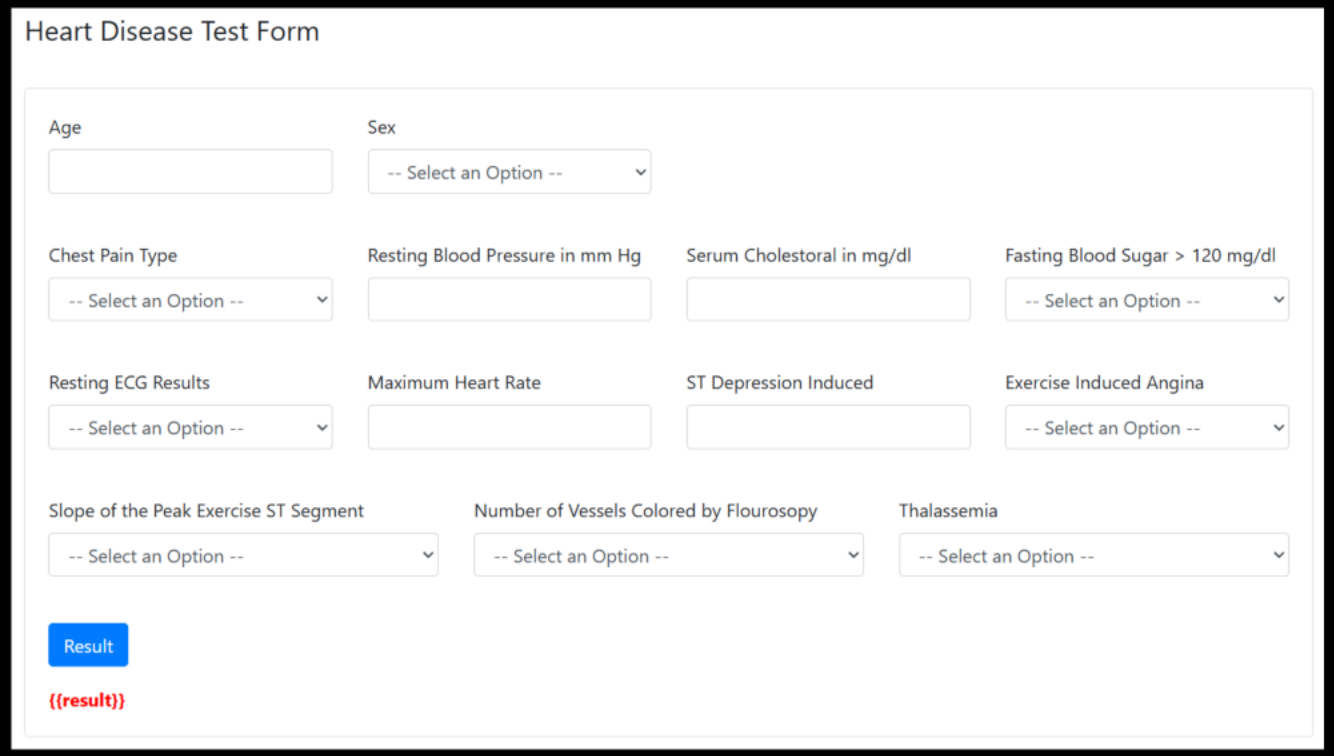


Accuracy measures the overall correctness of the model's predictions, precision focuses on the ratio of true positives to all predicted positives, and recall assesses the ability to identify true positives from all actual positives.

ROC Curve and AUC



The Receiver Operating Characteristic (ROC) curve visualizes the trade-off between true positive rate and false positive rate, while the Area Under the Curve (AUC) quantifies the model's discrimination ability. A higher AUC indicates better model performance.



WEB APPLICATION

MOTIVE

01 The web page makes heart disease prediction accessible to users instantaneously, enabling easy access without software installation.

02 It educates users about heart disease risk factors and promotes proactive health monitoring through timely predictions.

03 A web-based prediction tool allows for quick and timely assessments. Users can input their data, receive instant feedback on their potential risk of heart disease, and take proactive measures if necessary.

Languages Utilised

For a web-based heart disease prediction application, the choice of programming languages typically revolves around HTML, CSS, JavaScript and Python.

Python's Flask API is utilised to connect the trained machine learning model to the Web Page.

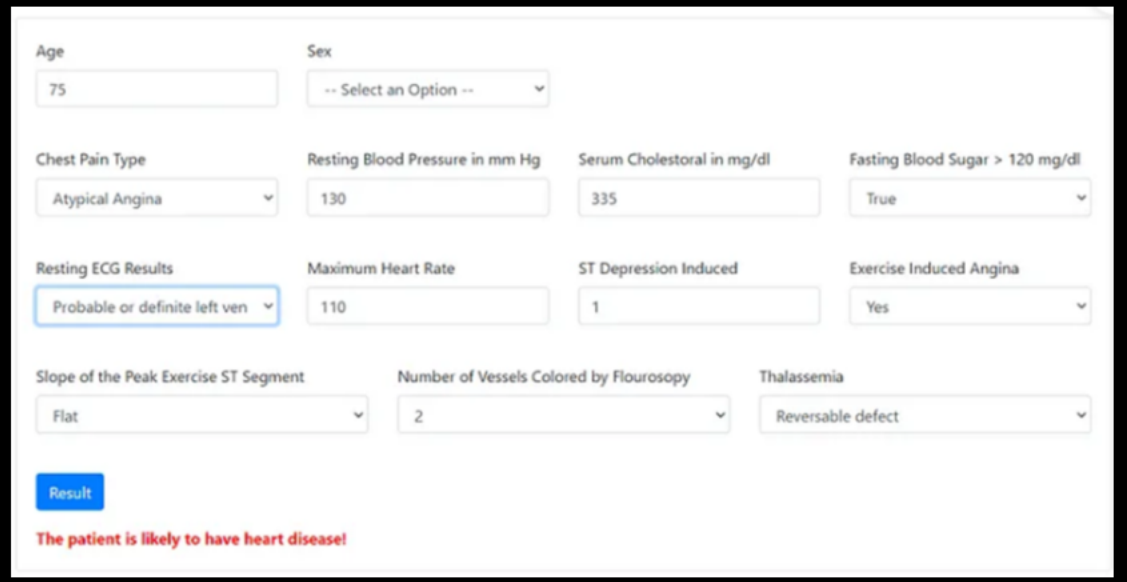
We use HTML and CSS for web page designing and customization. JavaScript is used to make elements of the web page more responsive.

Case Study

Exploring a Real-World Application of Machine Learning in Predicting Heart Disease

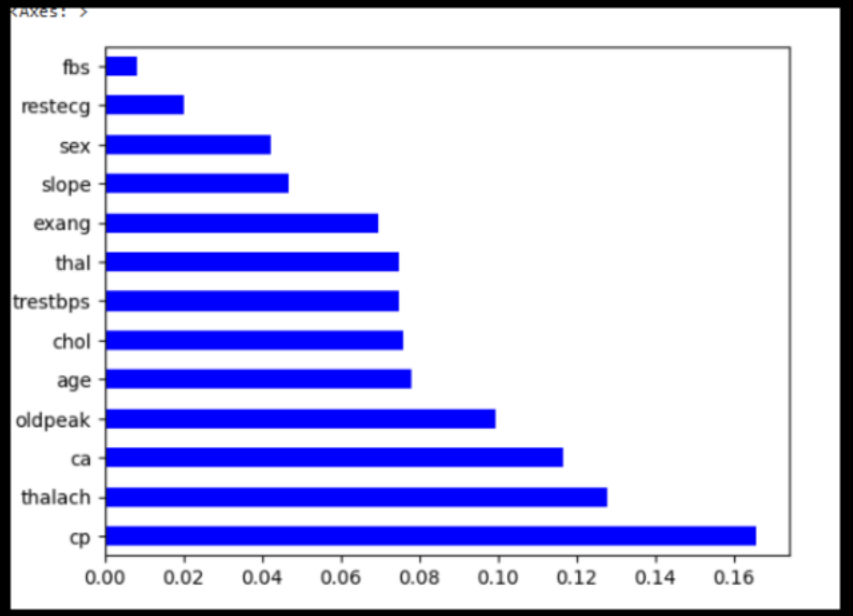
Live Implementation of Features

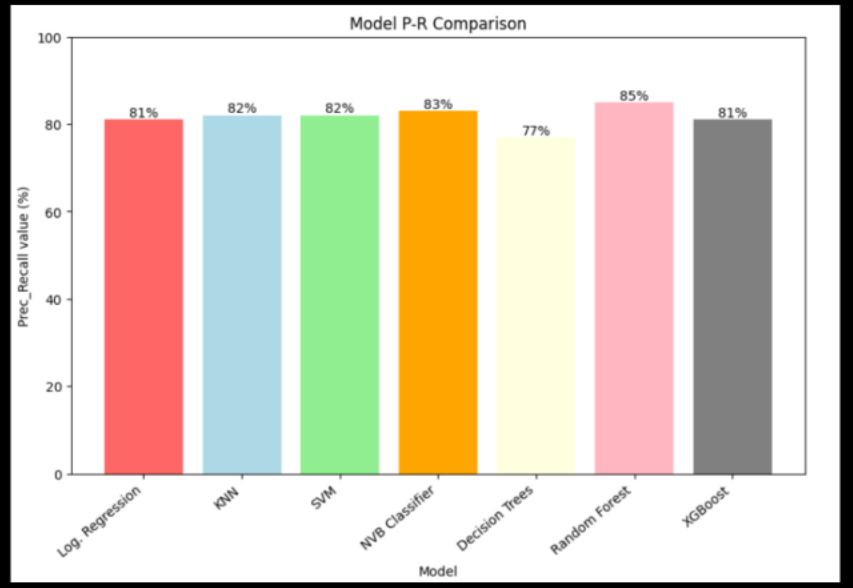
In the case study, we will apply real world values to test and observe the predictions made by the Web Page using our Trained Machine Learning Model.



Results and Insights

The results of the study revealed accurate predictions of heart disease with significant insights into the key factors contributing to cardiac health and potential risks for patients.





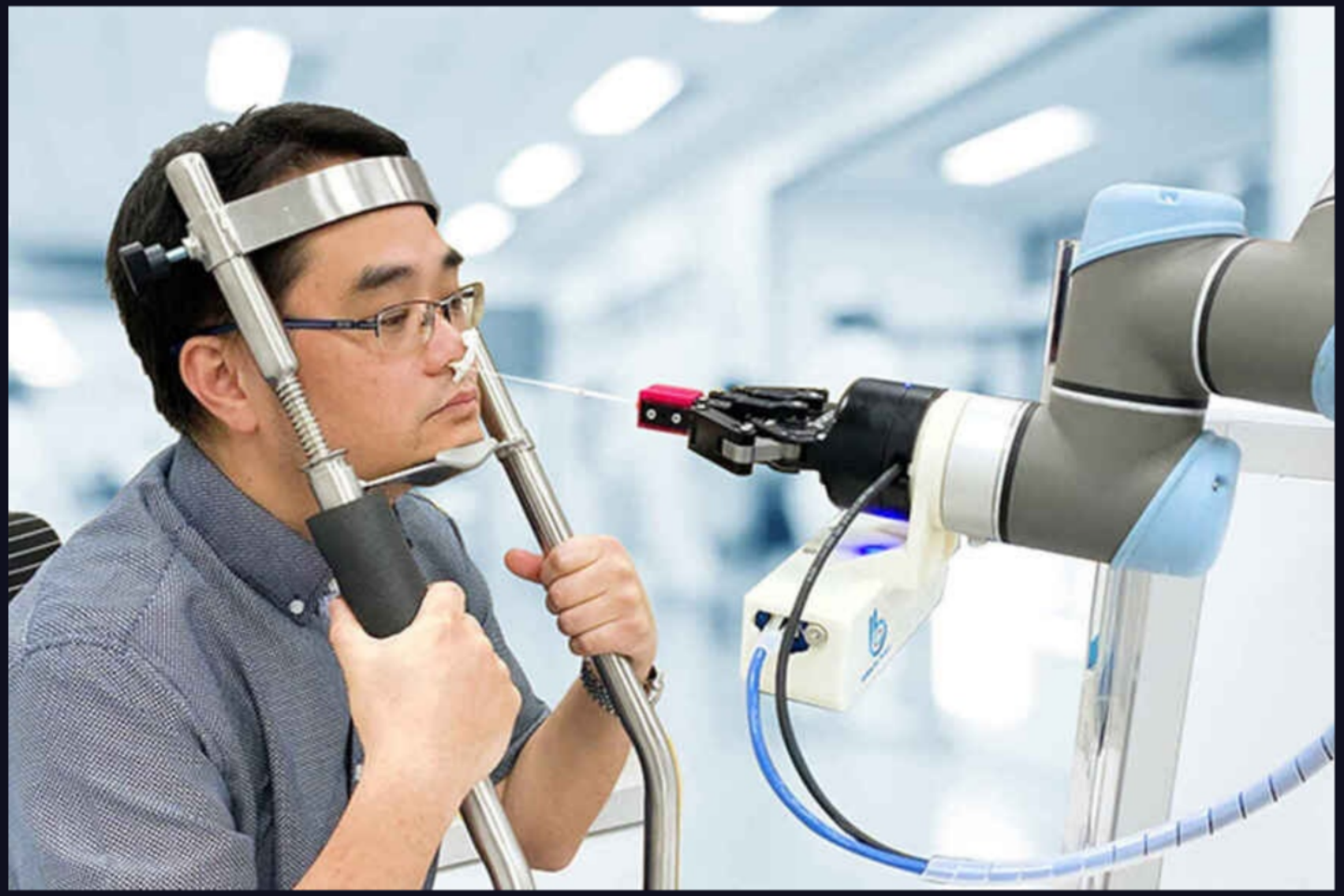
CONCLUSION
&
FUTURE SCOPE

Impact on Healthcare Industry

For many years to come Machine Learning and its relative technological implementations and inventions will result in drastic improvement in the Healthcare Industry.

Diagnosis, Treatments, Preventive Measures wil observe leaps and bounds of improvements as well as the Health Care industry involves technology to their advantage.

Future Research Directions



Further research in machine learning for heart disease prediction shall aim to enhance model accuracy, interpretability, and scalability.

Improved feature selection techniques and ensemble methods could lead to more reliable predictive models that support clinical decision-making.

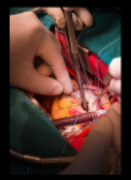

Thank you!

CARDIOVASCULAR RISK ASSESSMENT

ETE | JULY 2024
GUIDE: DR. DINESH SHARMA
SCHOOL OF INFORMATION AND TECHNOLOGY

Problem Statement

Overview of Heart Disease



Understanding Heart Disease

Heart Disease refers to the various conditions that affect the heart's functioning and structure.

It is one of the leading causes of mortality worldwide and needs to be addressed immediately.

Understanding Heart Disease involves:

- recognizing its types
- factors
- symptoms
- diagnostic methods
- treatment options
- preventive measures.

CONCLUSION
&
FUTURE SCOPE

Machine Learning Techniques

Importance of Data in Predictive Modeling

WEB APPLICATION

Case Study

Exploring a Real-World Application of Machine Learning in Predicting Heart Disease