Title: Prediction of heart failure using IBM Auto AI

Contents

Prediction of HeartFailure using IBM Auto AI

Introduction:

Heart failure (HF) is a complex clinical syndrome and not a disease. It prevents the heart from fulfilling the circulatory demands of the body, since it impairs the ability of the ventricle to fill or eject blood. It is characterized by symptoms, such as breathlessness, ankle swelling and fatigue that may be accompanied by signs, pulmonary crackles, and caused by structural and/or functional cardiac or non-cardiac abnormalities. HF is a serious condition associated with high morbidity and mortality rates. According to the European Society of Cardiology (ESC), 26 million adults globally are diagnosed with HF, while 3.6 million are newly diagnosed every year. 17–45% of the patients suffering from HF die within the first year and the remaining die within 5 years. The related to HF management costs are approximately 1–2% of all healthcare expenditure.

Over view:

Cardiovascular diseases (CVDs) are the number 1 cause of death globally, taking an estimated 17.9 million lives each year, which accounts for 31% of all deaths worldwide.

Heart failure is a common event caused by CVDs and this dataset contains 9 features that can be used to predict mortality by heart failure.
In this project, you need to build a model using Auto AI and build a web application where we can showcase the prediction of heart failure.

Purpose:

This project aims to explore the use of machine learning algorithms to predict Heart Failure. By using some of the significant features in our dataset and Using Software like IBM Auto AI and Machine Algorithms like, Random forest classifier, we can try to predict the premium costs of an insurance policy.

Existing problem:

The most crucial problem associated with what exists today is the

Heart failure.

Heart disease develops when there is:

- damage to all or part of the heart
- a problem with the blood vessels leading to or from the heart
- a low supply of oxygen and nutrients to the heart

In some cases, there is a genetic cause. However, some lifestyle factors and medical conditions can also increase the risk. These include:

- age
- smoking
- family history
- cholesterol
- BMI and some other factors.

Proposed solution:

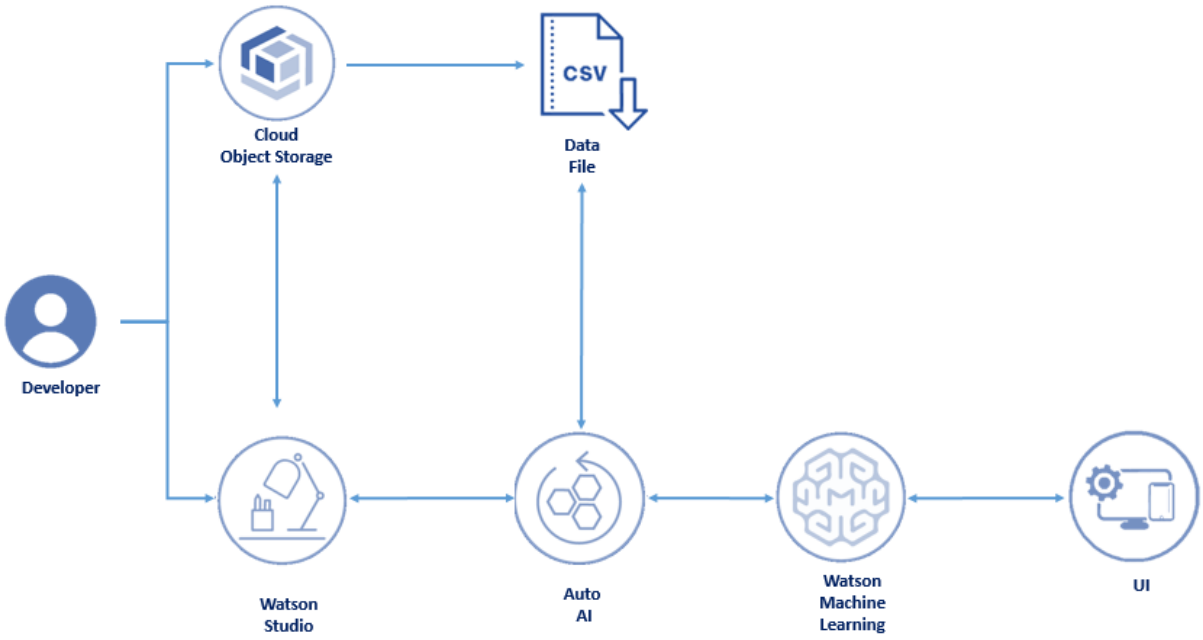
In this application, we study the effects of age, smoking, BMI, sex,

cholesterol, heartbeat per min, palpitations per min to determine how much of a difference these factors can make us to predict whether a person is going to get heart failure or not. This project aims at building a web App that automatically estimates prediction by taking the input values. By using our application, heart failure prediction becomes easy. The following are the methodologies adopted for data analysis, prediction and display of results by integrating with the system:

- IBM Auto AI
- Node Red application

Theoretical Analysis

Block Diagram



a model from a dataset that includes the age, gender, BMI, sex, family history, smoking preferences, family history to predict the heart failure. Using IBM Auto AI and Machine Learning algorithms, we automate all of the tasks involved in building predictive models for different requirements. Node Red is a framework that helps build web apps that could act as the interface to the user for input and output in the front and

integration of values in the backend.

Services Used:

- IBM Watson Studio
- IBM Watson Machine Learning
- Node-RED
- IBM Cloud Object Storage

Technical Architecture:

The user creates an IBM Watson Studio Service, IBM Cloud Object Storage Service on IBM Cloud.

● the user uploads the insurance premium data file into Watson Studio.
● the user creates an Auto AI Experiment to predict an insurance premium on Watson Studio.
● Auto AI uses Machine Learning services to create several models, and the user deploys the best performing model.

Hardware / Software designing:

Hardware: ● Lenovo Ideapad-500/8Gb RAM/64 bit/Windows 10.

Software: ●IBM Cloud

Experimental Investigations:

Dataset: SmartInternz provided the data source link, and the source of data for this project was from Kaggle. The dataset is comprised of 10801 records with 10 attributes. Attributes are as follow age, sex, bmi, smoker,

palpitationsper day and cholesterol, heartbeatspermin, familyhistory. The data was in a structured format and was stores in a CSV file. In a dataset, not every attribute has an impact on the prediction. Whereas some attributes even decline the accuracy, so it becomes necessary to remove these attributes from the features of the code. Removing such attributes not only help in improving accuracy but also the overall performance and speed.
Machine Learning can be defined as the process of teaching a computer system that allows it to make accurate predictions after the data is fed.

Binary Classification:

It is a type of classification with two outcomes, for e.g. – either true or false.
There are quite a few different algorithms used in binary classification In that Random Forest classifier, etc.
We have used Random Forest classifier for this model.

Random Forest Classifier:

Random forest is a supervised learning algorithm which is used for both classification as well as regression. But however, it is mainly used for classification problems. As we know that a forest is made up of trees and more trees means more robust forest. Similarly, random forest algorithm creates decision trees on data samples and then gets the prediction from each of them and finally selects the best solution by means of voting. It is an ensemble method which is better than a single decision tree because it reduces the over-fitting by averaging the result.

Result:

• Prediction using IBM Auto AI:

With Random forest classifier model using age, sex, BMI, palpitationperday, avgheartbeat, smoker, familyhistory, cholesterol in “Pipeline 3” is the top performer and gives accuracy of 86%. So the model or Pipeline 3 gives you more accuracy in lesstime, which is the best model by the prediction using IBM Auto AI.

ADVANTAGES:

1. Increased accuracy for effective heart disease diagnosis.
2. Handles roughest (enormous) amount of data using random forest algorithm and feature selection.
3. Reduce the time complexity of doctors.
4. Cost effective for patients.

DISADVANTAGES:

1.Prediction of cardiovascular disease results is not accurate.

2. Data mining techniques does not help to provide effective decision making.

3. Cannot handle enormous datasets for patient records.

Applications:

- Applications that help doctors or the patient directly suggesting

  Preventative healthy behaviours and habits to patients.

- Applications that used by dieticians, instructors or individuals

  Healthcare expenditures.

Conclusion:

In our data analysis we have used IBM Auto AI and Random Forest Classifier, Decision tree classifier for Prediction using Machine Learning Models to evaluate prediction of heart failure data.

Prediction using IBM Auto AI:

It has been found that the Random Forest Classifier model, “Pipeline 3” as the top performer with accuracy 86%. So the model which gives you in less time with more accuracy that will be taken into consideration.

Future Scope:

The scope of the heart failure prediction can be extended by adding new features to the model and saving space and memory used to build the model. The application could also include text facility for the users to give details about the heart failure. New Machine Learning techniques can we used to predict the model easily and calculate the predict value. It may also allow software 14 applications to become accurate in predicting outcomes. It enables machines to make data-driven decisions, which is more efficient than explicitly programming to carry out certain task.

Bibliography:

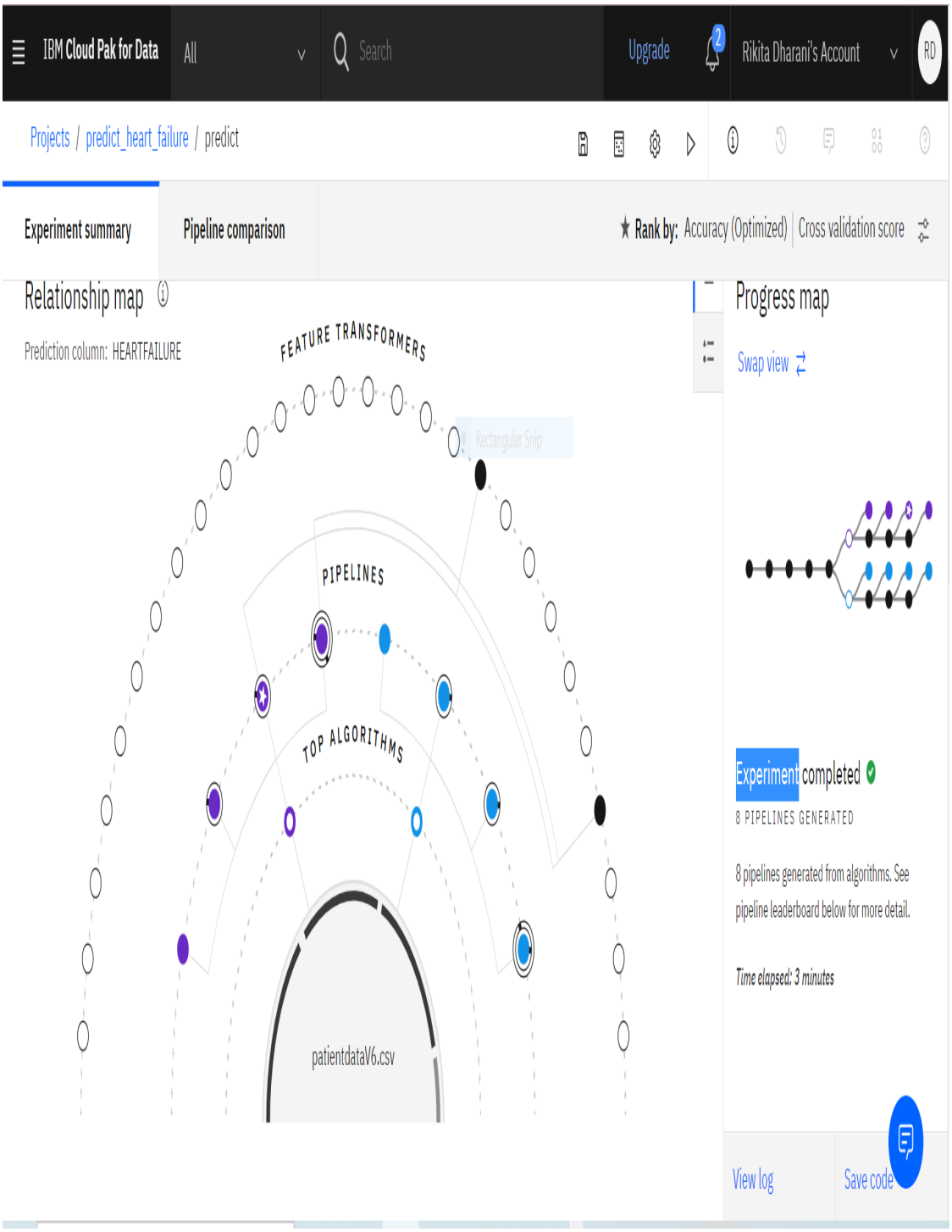
1. C. A. Devi, S. P. Rajamhoana, K. Umamaheswari, R. Kiruba, K. Karunya, and R. Deepika, "Analysis of neural networks based heart disease prediction system," in Proc. 11th Int. Conf. Hum. Syst. Interact. (HSI), Gdansk, Poland, Jul. 2018, pp. 233–239.

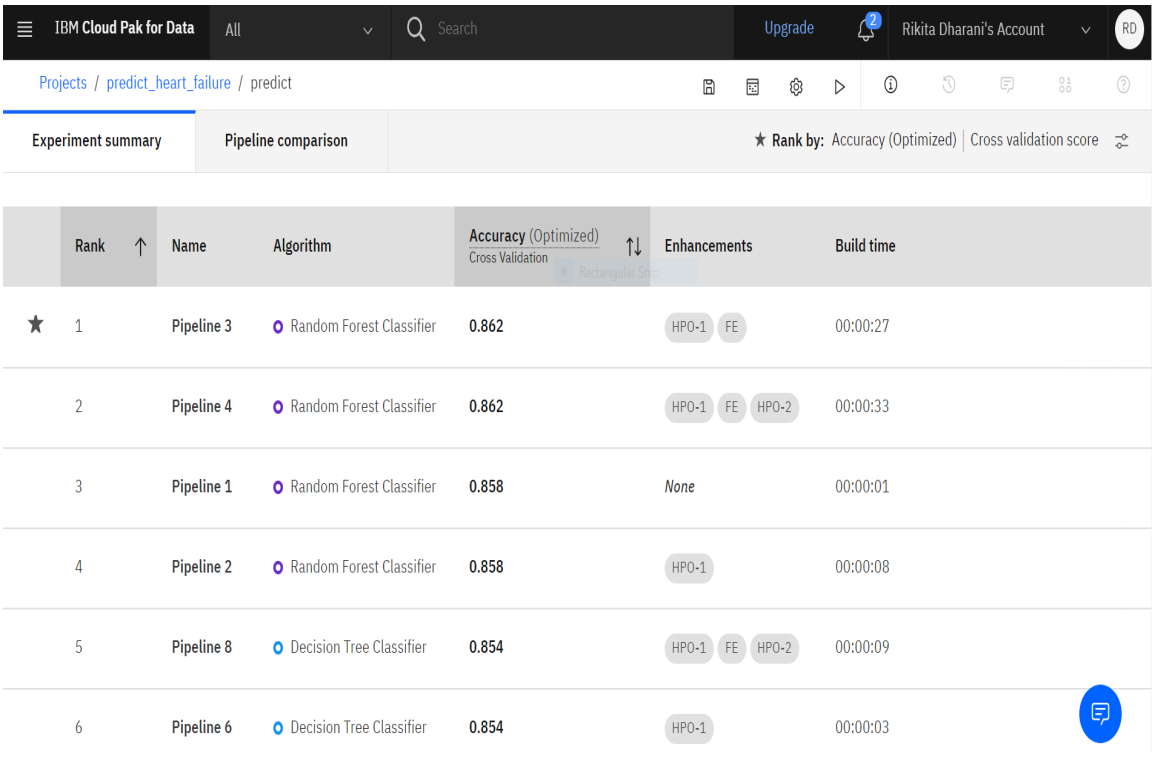
2. P. K. Anooj, "Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules," J. King Saud Univ.- Comput. Inf. Sci., vol.24, no.1, pp.27– 40,Jan.2012.doi:10.1016/j.jksuci.2011.09.002.

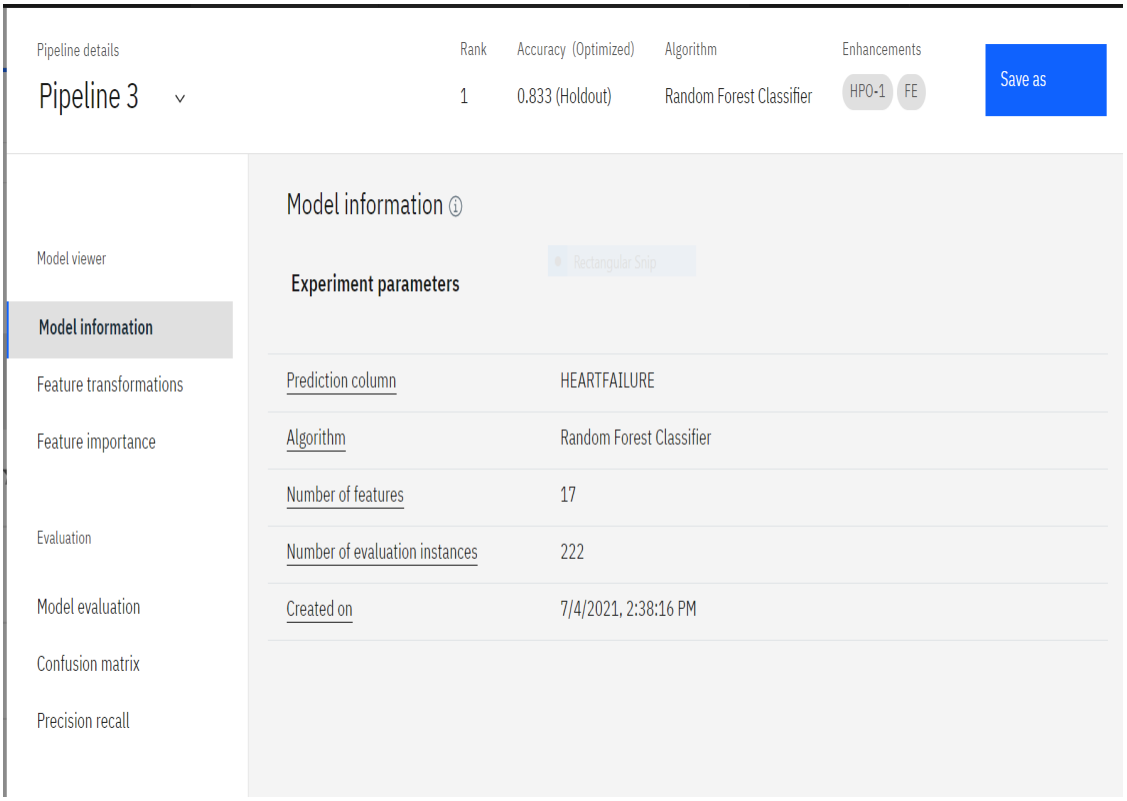
Appendix:

a. Source code

- **Prediction using IBM Auto AI:**


IBM Cloud Pak for Data
All
Search
Upgrade
2
Rikita Dharani's Account
RD
Projects / predict_heart_failure / predict
Experiment summary
Pipeline comparison
Rank by: Accuracy (Optimized) | Cross validation score
Relationship map
Prediction column: HEARTFAILURE
FEATURE TRANSFORMERS
PIPELINES
TOP ALGORITHMS
patientdataV6.csv
Progress map
Swap view
Experiment completed
8 PIPELINES GENERATED
8 pipelines generated from algorithms. See pipeline leaderboard below for more detail.
Time elapsed: 3 minutes
View log
Save code


IBM Cloud Pak for Data
All
Search
Upgrade
Rikita Dharani's Account
RD
Projects / predict_heart_failure / predict
Experiment summary
Pipeline comparison
Rank by: Accuracy (Optimized) | Cross validation score
Rank
Name
Algorithm
Accuracy (Optimized)
Cross Validation
Enhancements
Build time
1 Pipeline 3 Random Forest Classifier 0.862 HPO-1 FE 00:00:27
2 Pipeline 4 Random Forest Classifier 0.862 HPO-1 FE HPO-2 00:00:33
3 Pipeline 1 Random Forest Classifier 0.858 None 00:00:01
4 Pipeline 2 Random Forest Classifier 0.858 HPO-1 00:00:08
5 Pipeline 8 Decision Tree Classifier 0.854 HPO-1 FE HPO-2 00:00:09
6 Pipeline 6 Decision Tree Classifier 0.854 HPO-1 00:00:03


Pipeline details
Pipeline 3
Rank
1
Accuracy (Optimized)
0.833 (Holdout)
Algorithm
Random Forest Classifier
Enhancements
HPO-1 FE
Save as
Model viewer
Model information
Feature transformations
Feature importance
Evaluation
Model evaluation
Confusion matrix
Precision recall
Model information
Experiment parameters
Prediction column HEARTFAILURE
Algorithm Random Forest Classifier
Number of features 17
Number of evaluation instances 222
Created on 7/4/2021, 2:38:16 PM

Pipeline details

Pipeline 3

Rank	Accuracy (Optimized)	Algorithm	Enhancements
1	0.833 (Holdout)	Random Forest Classifier	HPO-1 FE

Save as

X

Model viewer

Model information

Feature transformations

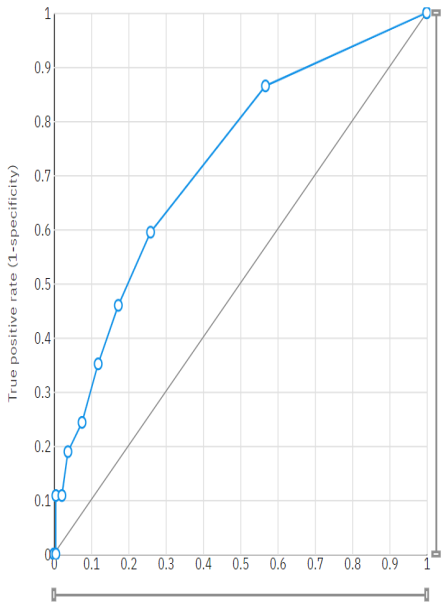
Feature importance

Evaluation

Model evaluation

Confusion matrix

Precision recall



Pipeline details

Pipeline 3

Rank	Accuracy (Optimized)	Algorithm	Enhancements
1	0.833 (Holdout)	Random Forest Classifier	HPO-1 FE

Save as

Model viewer

Model information

Feature transformations

Feature importance

Evaluation

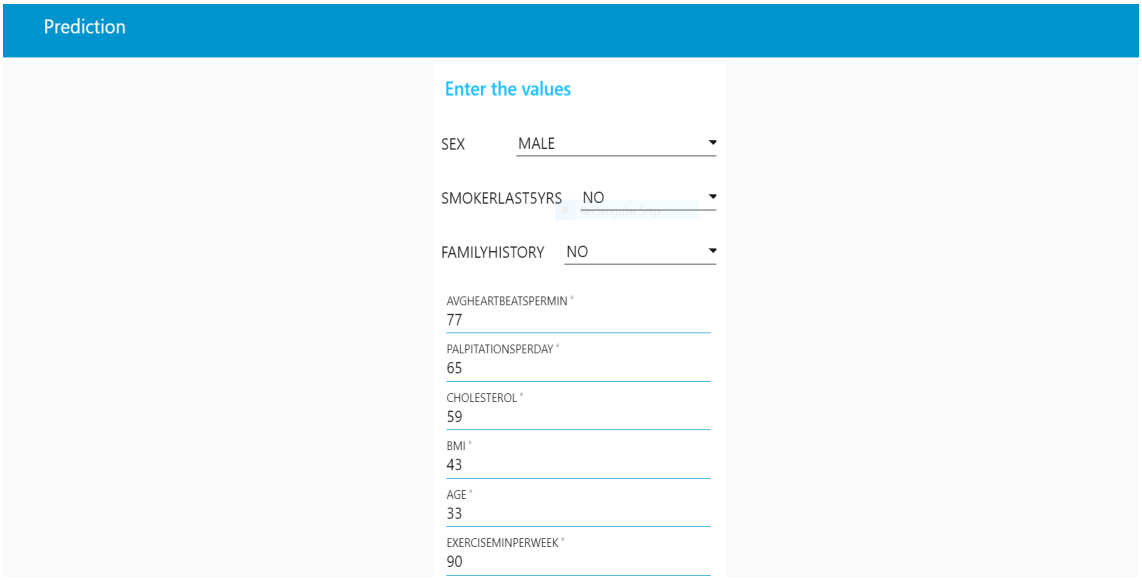
Model evaluation

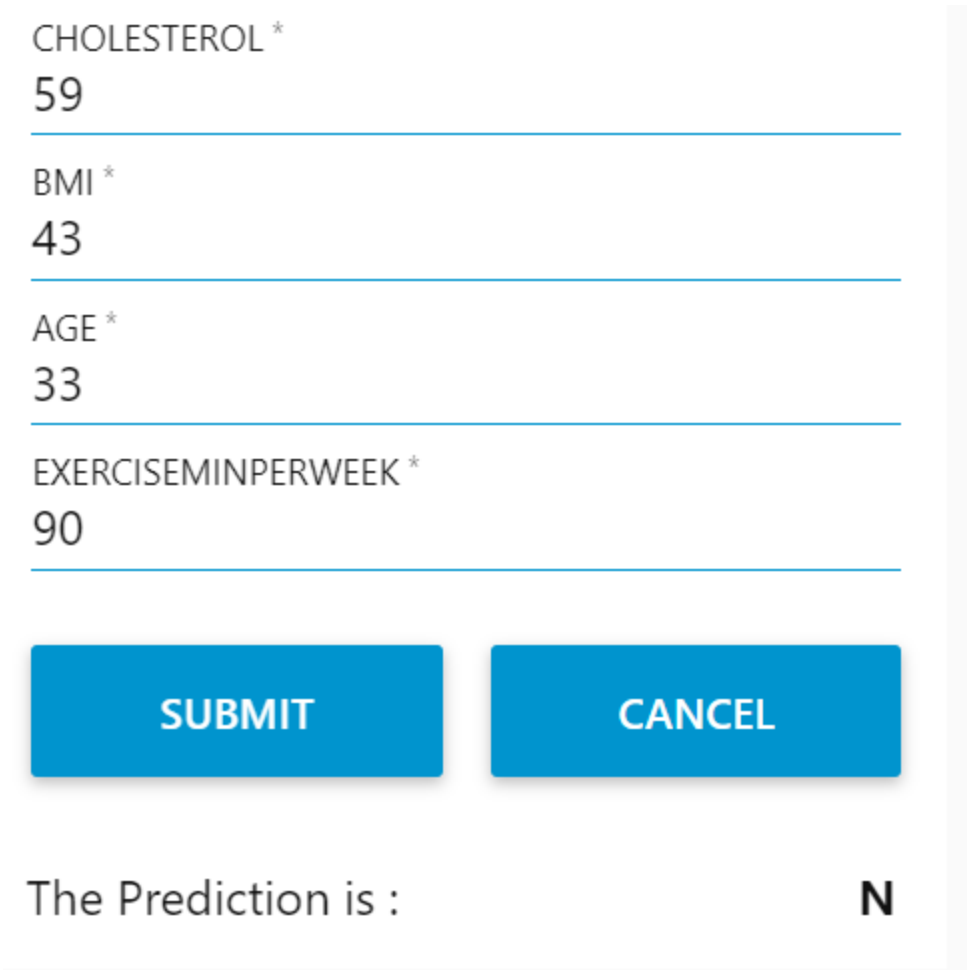
Confusion matrix

Precision recall

Model evaluation measure

Measures	Holdout score	Cross validation score
Accuracy	0.833	0.862
Area under ROC	0.721	0.749
Precision	0.500	0.684
Recall	0.189	0.311
F1	0.275	0.427
Average precision	0.354	0.462
Log loss	1.112	1.319


Prediction
Enter the values
SEX
MALE
SMOKERLAST5YRS
NO
FAMILYHISTORY
NO
AVGHEARTBEATSPERMIN *
77
PALPITATIONSPERDAY *
65
CHOLESTEROL *
59
BMI *
43
AGE *
33
EXERCISEMINPERWEEK *
90


CHOLESTEROL *
59
BMI *
43
AGE *
33
EXERCISEMINPERWEEK *
90
SUBMIT
CANCEL
The Prediction is :
N