

1. Analysis of results:

Projects / Effective Heart Disease Predictio... / Heart Disease Prediction						
Experiment summary		Pipeline comparison		★ Rank by: Accuracy (Optimized) Cross validation score		
★						
1	Pipeline 3	XGB Classifier	0.873	HPO-1 FE	00:00:34	
2	Pipeline 4	XGB Classifier	0.873	HPO-1 FE HPO-2	00:01:16	
3	Pipeline 1	XGB Classifier	0.869	None	00:00:01	
4	Pipeline 2	XGB Classifier	0.869	HPO-1	00:00:09	
5	Pipeline 7	Random Forest Classifier	0.862	HPO-1 FE	00:00:37	
6	Pipeline 8	Random Forest Classifier	0.862	HPO-1 FE HPO-2	00:01:14	
7	Pipeline 5	Random Forest Classifier	0.858	None	00:00:01	
8	Pipeline 6	Random Forest Classifier	0.858	HPO-1	00:00:10	

Figure -1: Performance of different machine learning techniques

It has been observed that, XGB Classifier outperformed as compare to Random Forest Classifier with accuracy of 87.3% within response time of 34 sec and hence it has been choose for deployment. Random Forest on the other hand gave its best accuracy at 86.2%. The ROC, precision and recall of XGB classifier is given as follows.

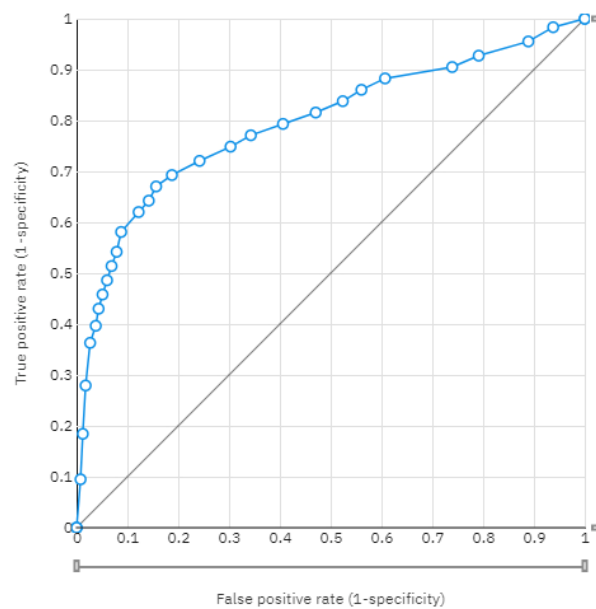


Figure -2: ROC of XGB Classifier

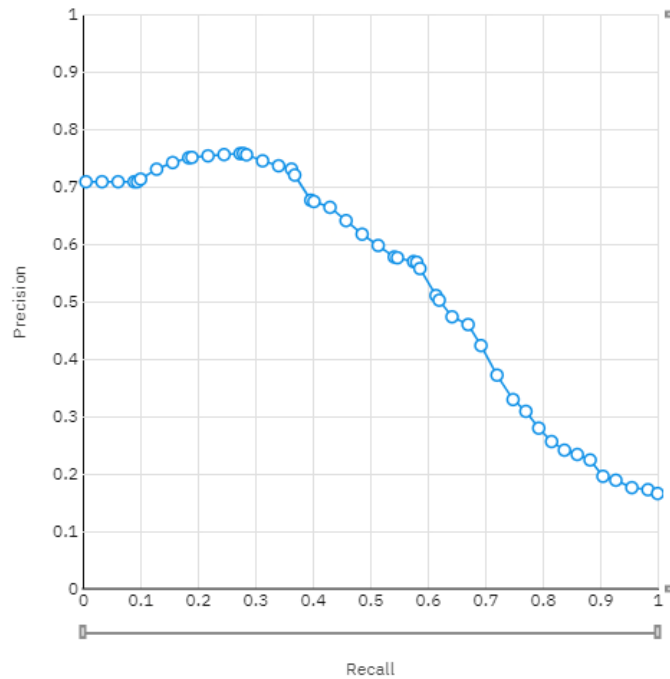


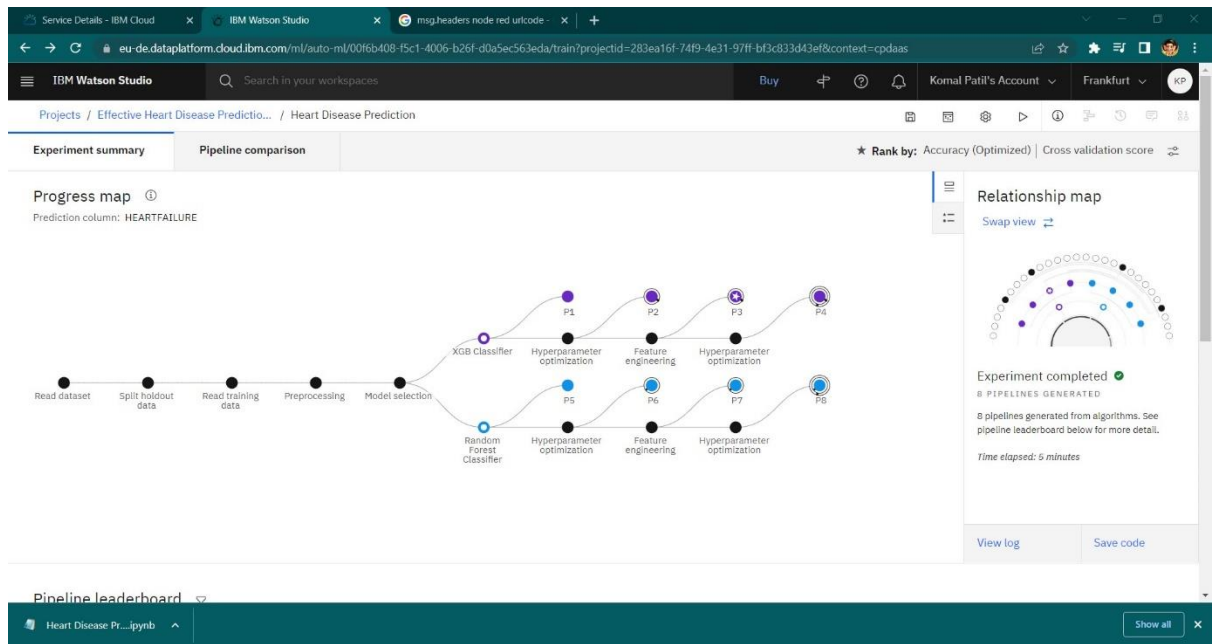
Figure -3: Precision and Recall of XGB Classifier

2. Outputs:

The feature summary of the dataset is given below

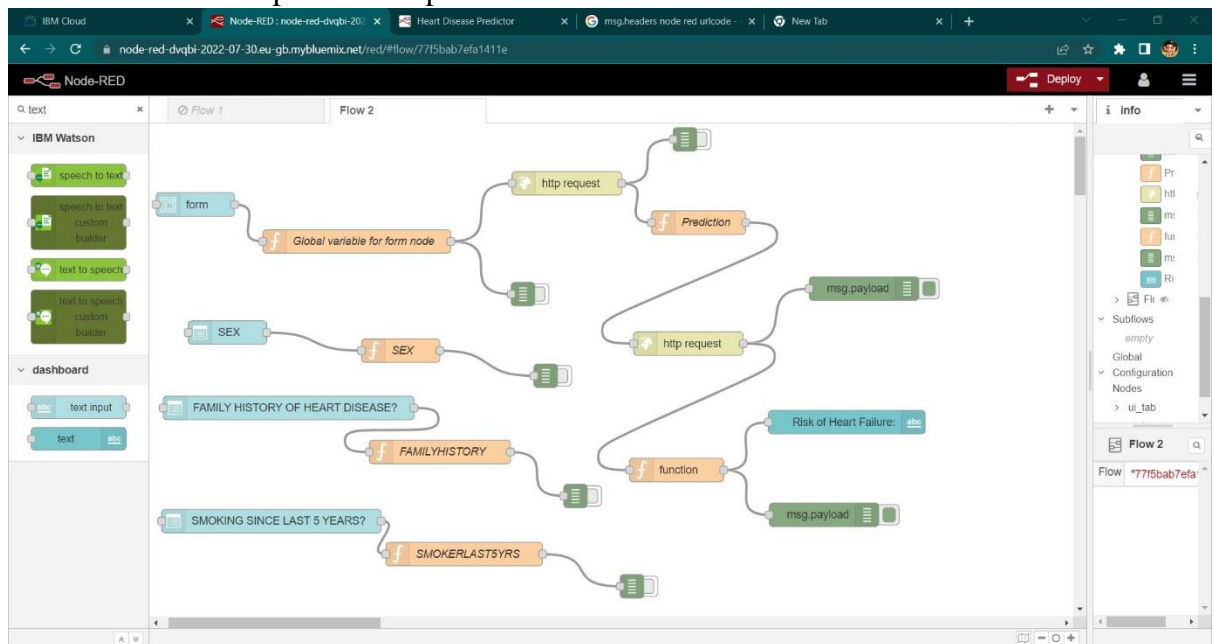
Pipeline details			Rank	Accuracy (Optimized)	Algorithm	Enhancements	Save as
Pipeline 3			1	0.869 (Holdout)	XGB Classifier	HPO-1 FE	
Model viewer	FAMILYHISTORY	None		100.00%			
Model information	BMI	None		12.00%			
Feature summary	SMOKERLASTSYRS	None		8.00%			
Evaluation	PALPITATIONSPERDAY	None		7.00%			
Model evaluation	EXERCISEMINPERWEEK	None		6.00%			
Confusion matrix	AVGHEARTBEATSPERMIN	None		6.00%			
Precision recall	CHOLESTEROL	None		4.00%			
	✓ NewFeature_0	sum(AVGHEARTBEATSPERMIN, PALPITATIONSPERDAY)		4.00%			
	✓ NewFeature_2	sum(AVGHEARTBEATSPERMIN, AGE)		3.00%			
	✓ NewFeature_5	sum(PALPITATIONSPERDAY, EXERCISEMINPERWEEK)		1.00%			

According to the feature summary, family history of heart disease is the most significant attribute for classification followed by BMI, Smoking habit, Palpitation per day, Exercise per week (in min), average heartbeat and cholesterol.



In the given pipeline, Pipeline3(XGB Classifier) gave the highest accuracy and chosen for deployment.

The following Node-RED flow has been designed for UI application; it is integrated with the trained model to provide the predictions.



The final Web application of the project is given in the following figure

IBM CloudNode-RED: node-red-dvqbi-20Heart Disease Predictormsg.headers node red urlcodeNew Tabnode-red-dvqbi-2022-07-30.eu-gb.mybluemix.net/u/#/0?socketid=zaDtoIag7h6bs1IAABF

Heart Disease Predictor

HEART DISEASE PREDICTOR

SEX M

FAMILY HISTORY OF HEART DISEASE? N

SMOKING SINCE LAST 5 YEARS? N

AVERAGE HEARTBEATS(per minute) ¹
93

PALPITATIONS(PERDAY) ¹
22

CHOLESTEROL ¹
150

BMI ¹
24

AGE ¹
45

EXERCISE PER WEEK(in minutes) ¹
110

SUBMITCANCEL

Risk of Heart Failure: **N**