Heart Attack Risk Analysis & Prediction Using IBM Cloud's Watson Studio and AutoAI

INTRODUCTION:

Overview

Heart disease is related to all diverse diseases affecting the heart. The exposure of heart disease from various factors is an issue which is not free from false presumptions often accompanied by unpredictable effects. Researchers have been using several data mining techniques to help health care professionals in the diagnosis of heart disease. In our work, proposed to determine the attributes which contribute more towards the diagnosis of heart ailments, which indirectly may reduce the number of tests which are needed to be taken by a patient. The dataset used consists of 270 cases. Originally, thirteen attributes were involved (age, sex, chest pain type, resting blood pressure, serum cholesterol, fasting blood sugar, resting electrocardiographic results, maximum heart rate, exercise induced angina, depression induced by exercise relative to rest, slope, number of vessels colored by fluoroscopy, and exercise thallium scintigraphy defects). To evaluate the performance of the model the sensitivity, specificity, and accuracy are calculated. Comparison with other data mining techniques is presented. The simulation result obtained from the model enables to establish significant patterns and relationships between the medical factors and heart disease.

Today, medical services have come a long way to treat patients with various diseases. Today, diagnosing patients correctly and administering effective treatments have become quite a challenge. Poor clinical decisions may end inpatient's death which could not be tolerated by the hospital as it loses its reputation. The cost to treat a patient with a heart problem is quite high and not affordable by every patient. To achieve a correct and cost- effective treatment, computer-based information and/or decision support systems can be developed to do the task. Most hospitals today use

some sort of hospital information systems to manage their healthcare or patient data.

Purpose :

This project gives the understanding of the heart disease prediction

- It might have happened so many times that you or someone yours need doctors help immediately, but they are not available due to some reason.

- The Heart Disease Prediction application is an end user support and online consultation project.

- Here, we propose a web application that allows users to get instant guidance on their heart disease through an intelligent system online.

- The application is fed with various details and the heart disease associated with those details.

- The application allows user to share their heart related issues.

- It then processes user specific details to check for various illness that could be associated with it.

- Here we use some intelligent data mining techniques to guess the most accurate illness that could be associated with patient's details.

- Based on result, system automatically shows the result specific doctors for further treatment.

- The system allows user to view doctor's details.

- The system can be use in case of emergency for heart disease prediction.

LITERATURE SURVEY

Heart is the next major organ comparing to brain which has more priority in Human body. It pumps the blood and supplies to all organs of the whole body. Prediction of occurrences of heart diseases in medical field is significant work. Data analytics is useful for prediction from more information and it helps medical Centre to predict of various disease

There are numerous works has been done related to disease prediction systems using different data mining techniques and machine learning algorithms in medical centres.

K.Polaraju et al, [1] proposed Prediction of Heart Disease using Multiple Regression Model and it proves that Multiple Linear Regression is appropriate for predicting heart disease chance. The work is performed using training data set consists of 3000 instances with 13 different attributes which has mentioned earlier. The data set is divided into two parts that is 70% of the data are used for training and 30% used for testing. Based on the results, it is clear that the classification accuracy of Regression algorithm is better compared to other algorithms.

Marjia et al, [2] developed heart disease prediction using KStar, j48, SMO, and Bayes Net and Multilayer perception using WEKA software. Based on performance from different factor SMO and Bayes Net achieve optimum performance than KStar, Multilayer perception and J48 techniques using kfold cross validation. The accuracy performances achieved by those algorithms are still not satisfactory. Therefore, the accuracy's performance is improved more to give better decision to diagnosis disease.

S. Seema et al,[3] focuses on techniques that can predict chronic disease by mining the data containing in historical health records using Naïve Bayes, Decision tree, Support Vector Machine(SVM) and Artificial Neural Network(ANN). A comparative study is performed on classifiers to measure the better performance on an accurate rate. From this experiment, SVM gives highest accuracy rate, whereas for diabetes Naïve Bayes gives the highest accuracy.

Ashok Kumar Dwivedi et al, [4] recommended different algorithms like Naive Bayes, Classification Tree, KNN, Logistic Regression, SVM and ANN. The Logistic Regression gives better accuracy compared to other algorithms.

MeghaShahi et al, [5] suggested Heart Disease Prediction System using Data Mining Techniques. WEKA software used for automatic diagnosis of disease and to give qualities of services in healthcare centres. The paper used various algorithms like SVM, Naïve Bayes, Association rule, KNN, ANN, and Decision Tree. The paper recommended SVM is effective and provides more accuracy as compared with other data mining algorithms.

Chala Beyene et al, [6] recommended Prediction and Analysis the occurrence of Heart Disease Using Data Mining Techniques. The main objective is to predict the occurrence of heart disease for early automatic diagnosis of the disease within result in short time. The proposed methodology is also critical in healthcare organisation with experts that have no more knowledge and skill. It uses different medical attributes such as blood sugar and heart rate, age, sex are some of the attributes are included to identify if the person has heart disease or not. Analyses of dataset are computed using WEKA software.

R. Sharmila et al, [7] proposed to use non- linear classification algorithm for heart disease prediction. It is proposed to use bigdata tools such as Hadoop Distributed File System (HDFS), Mapreduce along with SVM for prediction of heart disease with

optimized attribute set. This work made an investigation on the use of different data mining techniques for predicting heart diseases. It suggests to use HDFS for storing large data in different nodes and executing the prediction algorithm using SVM in more than one node simultaneously using SVM. SVM is used in parallel fashion which yielded better computation time than sequential SVM.

Jayami Patel et al, [8] suggested heart disease prediction using data mining and machine learning algorithm. The goal of this study is to extract hidden patterns by applying data mining techniques. The best algorithm J48 based on UCI data has the highest accuracy rate compared to LMT.

Purushottam et al, [9] proposed an efficient heart disease prediction system using data mining. This system helps medical practitioner to make effective decision making based on the certain parameter. By testing and training phase a certain parameter, it provides 86.3% accuracy in testing phase and 87.3% in training phase.

Proposed solution:

In this project random forest Binary classifier is the solution which has been choosen to finding out the best solution for outcome of the prediction of the patient is in risk or safe zone with certain percentages is been shown over us Data analytics is useful for prediction from more information and it helps medical centre to predict of various disease. Huge amount of patient related data is maintained on monthly basis. The stored data can be useful for source of predicting the occurrence of future disease.

THEORITICAL ANALYSIS

Flow chart for the procedure/Block diagram
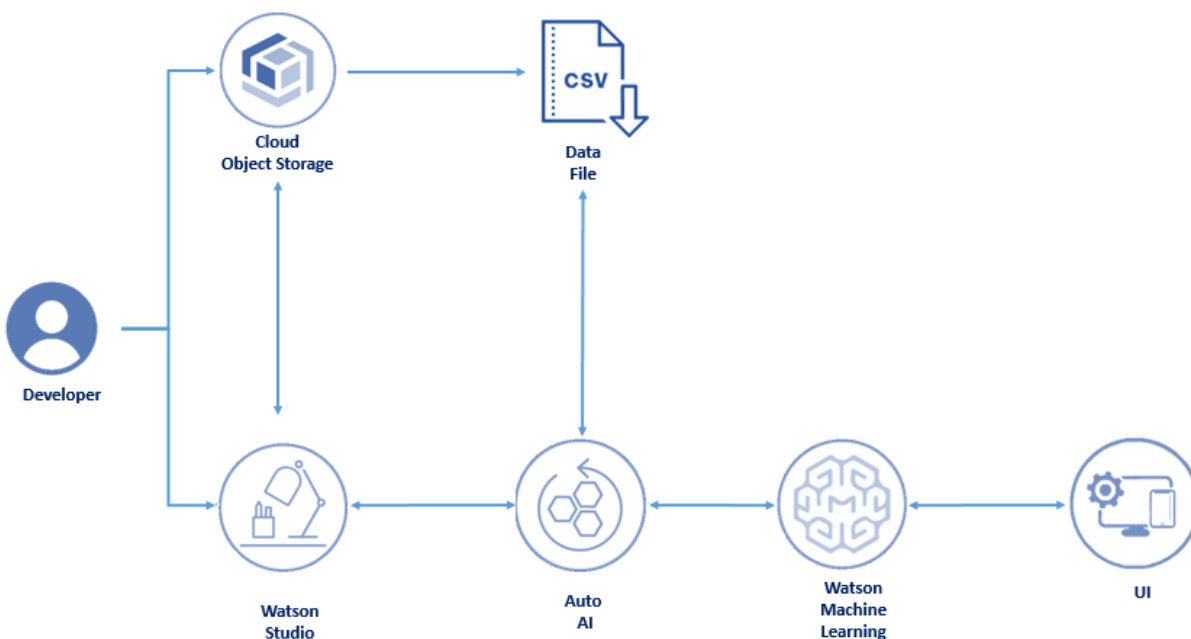
Hardware / Software designing:

Requirements:
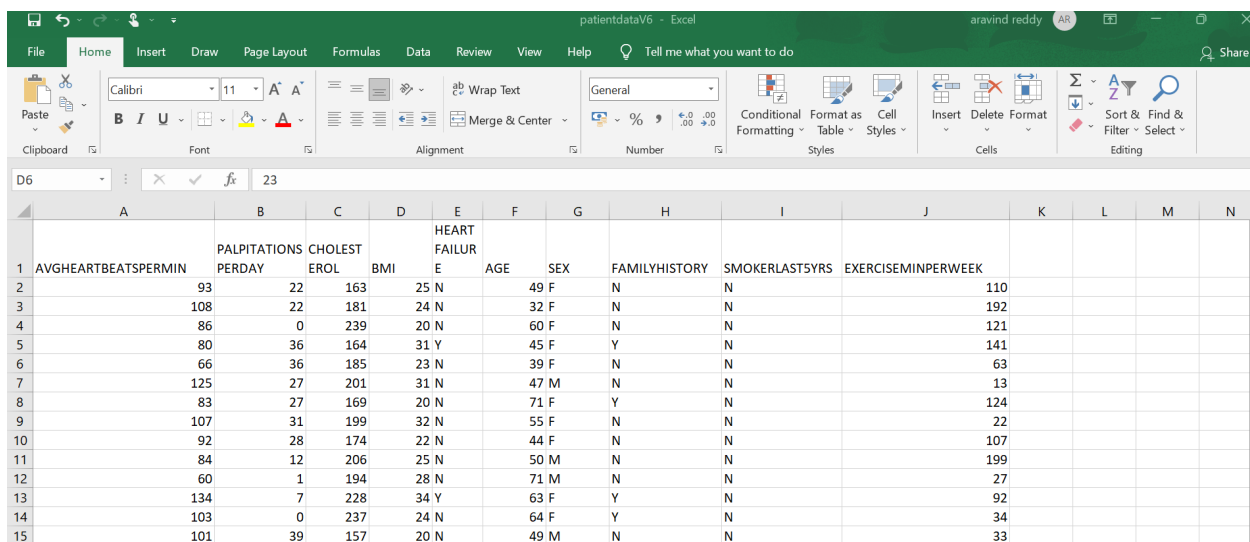
Laptop /Desktop

Windows Pc 10 4GB Ram at least

MS Office

Compactable browser for running IBM cloud such as Google chrome/Explorer

Skills Required:

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

EXPERIMENTAL INVESTIGATIONS:

For analysis the sample dataset which I have used stored in csv file and there fields definition are below

AVGHEARTBEATSPERMIN	PALPITATIONS PERDAY	CHOLESTEROL	BMI	HEART FAILURE	AGE	SEX	FAMILYHISTORY	SMOKERLAST5YRS	EXERCISEMINPERWEEK
93	22	163	25	N	49	F	N	N	110
108	22	181	24	N	32	F	N	N	192
86	0	239	20	N	60	F	N	N	121
80	36	164	31	Y	45	F	Y	N	141
66	36	185	23	N	39	F	N	N	63
125	27	201	31	N	47	M	N	N	13
83	27	169	20	N	71	F	Y	N	124
107	31	199	32	N	55	F	N	N	22
92	28	174	22	N	44	F	N	N	107
84	12	206	25	N	50	M	N	N	199
60	1	194	28	N	71	M	N	N	27
134	7	228	34	Y	63	F	Y	N	92
103	0	237	24	N	64	F	Y	N	34
101	39	157	20	N	49	M	N	N	33

In this Activity, we gonna build a machine learning model that predicts heart failure based on the following parameters

- AVGHEARTBEATSPERMIN

- PALPITATIONSPERDAY

- CHOLESTEROL

- BMI

- AGE

- SEX

- FAMILY HISTORY

- SMOKERLAST5YRS

- EXERCISEMINPERWEEK

A project is how you organize your resources to achieve a particular goal.

In this project we are finding the heart failure prediction using a random tree binary classifier algorithm.

There are the following steps for performing or deploying the machine learning model.

here some steps to follows. Log on to IBM Cloud Watson Studio and create new Project.
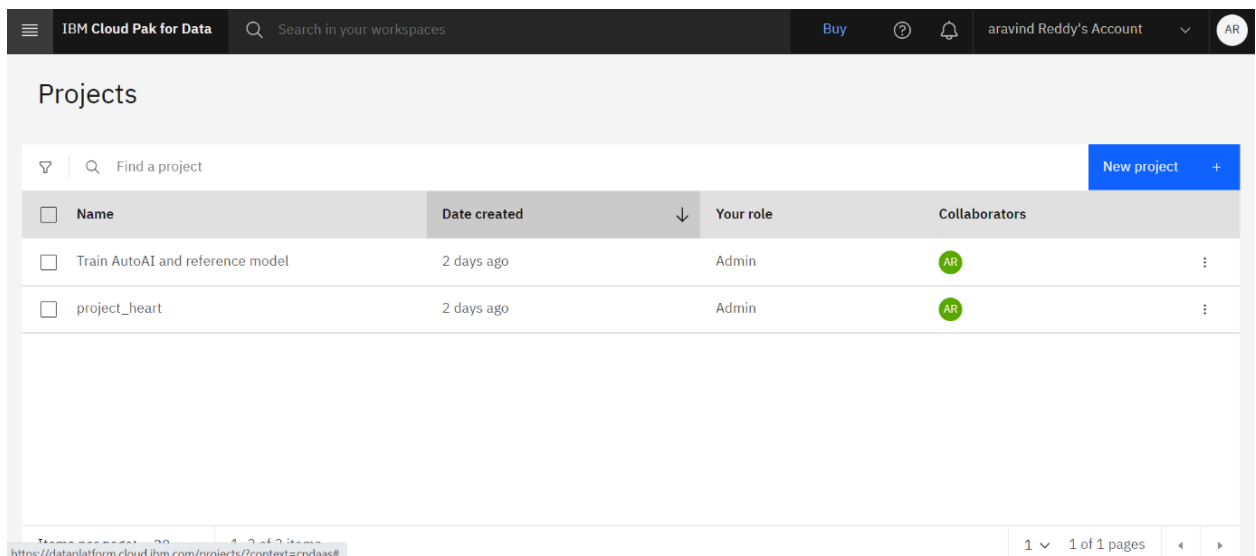
Before loading the dataset click on Auto AI services for performing or deploying the machine learning model.

After successfully created project add csv file by using "Add to project" button. and Choose Data Option
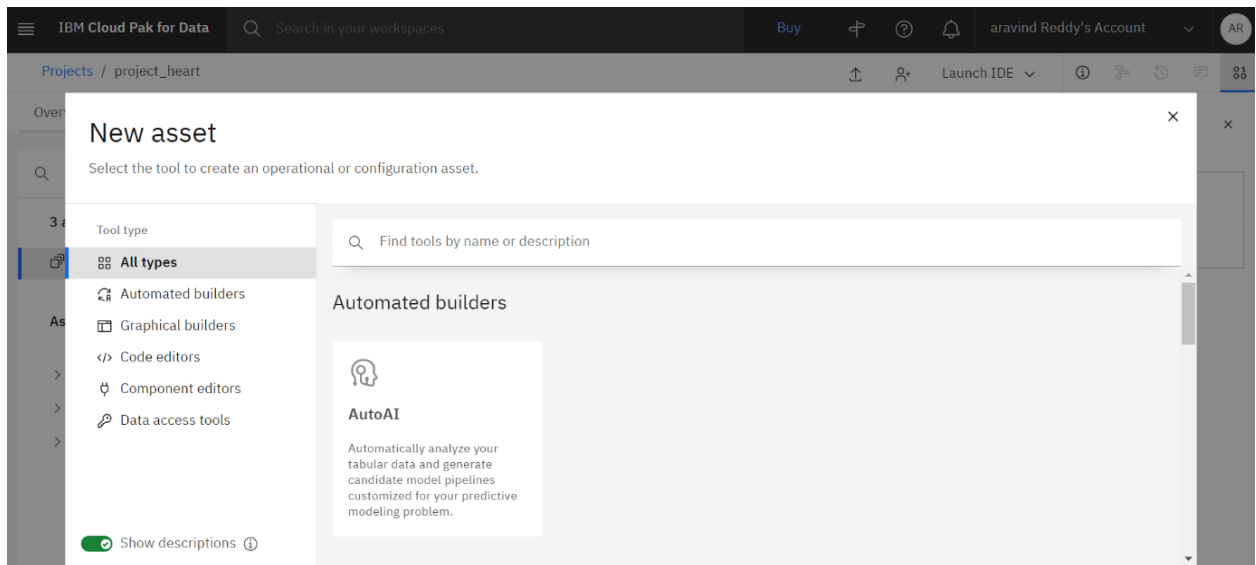
after successfully adding csv file dataset add AutoAI experiment. give it name and e.g "Heart Attack Model" and select predict column in my case output as predict column. and I choose Binary Classification for my AI model. then click on Run experiment button which will build your model.
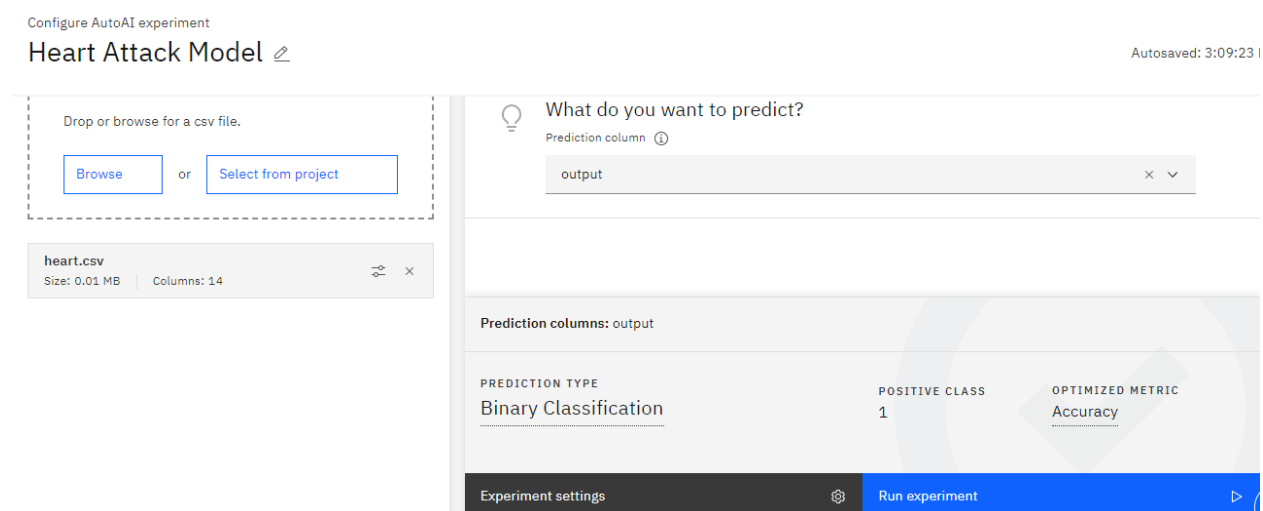
Auto AI build your model with details step and different algorithm. it also shows there different accuracy and categories. it also ranking best suitable algorithm for model. so its far easy to analysis data model using pipelines generated by AutoAI.In Relationship map shows all analysis from your data to feature transformers stepwise. You need to hover on using mouse it will show which algorithms used for build pipeline(model) in my case here are two top algorithms 1) Logistic Regression and 2) Extra Trees Classifier

Next level shows some pipelines with their accuracy and best suggested model as star marked. After pipelines next level shows used feature transformation to build pipeline.

When you click any of pipeline its will shows details analysis of how your model is build and evaluate. E.g Model Evaluation , Confusion Matrix , Precision Recall Curve.Model Evaluation : provides measure of overall predictive accuracy of the model. It also shows ROC Curve to show true positive rate and false positive rate on there axis , and Model Evaluations Measures table which have values about holdout Score and Cross Validation Score. Where we can detail analysis how its targeted accuracy measure.

Confusion Matrix : Shows numbers and proportions of correct and incorrect classifications of appropriate models. We can take decision which are less correct and more correct using percentages.Precision Recall Curve : this chart plats the proportion of outcomes predicted to be positive that are positive, on curve vertical axis against the proportion of positive outcomes that are correctly predicted.

In Model Viwer there are some more details about model features and evaluations. It also shows the relative importance of each feature in predicating the target, based on an averaging of different important measures or weight.Progress Map : it shows progress when your model building in process.

Pipeline Comparison : in this feature of analysis show matrix chart which compare the matrix from each pipeline and shows there score for each step of dragging prediction so you can change rank of it and change prediction by cross validation and holdout.
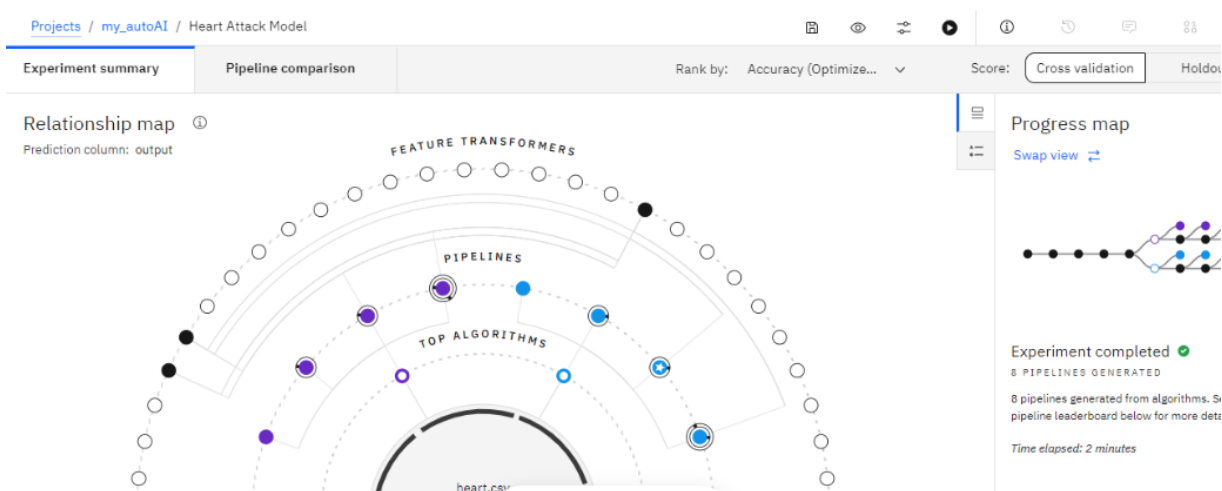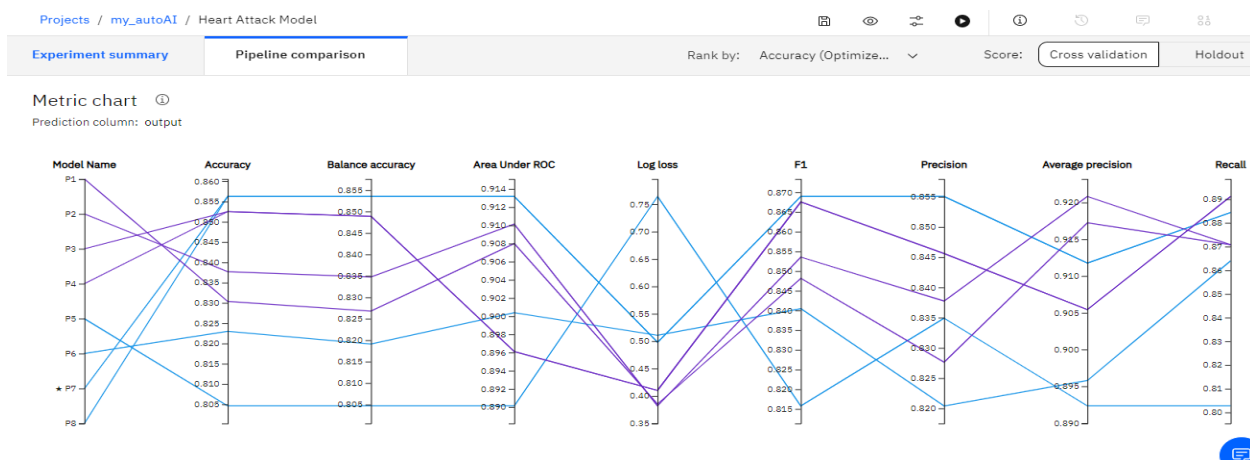
Experiment summary | Pipeline comparison

Rank by: Accuracy (Optimize... ⌄   Score: Cross validation | Holdou

Relationship map ⓘ

Prediction column: output

FEATURE TRANSFORMERS

PIPELINES

TOP ALGORITHMS

heart.csv

Progress map

Swap view ⇄

Experiment completed ✓

8 PIPELINES GENERATED

8 pipelines generated from algorithms. Se
pipeline leaderboard below for more deta

Time elapsed: 2 minutes

Experiment summary | Pipeline comparison

Rank by: Accuracy (Optimize... ⌄   Score: Cross validation | Hold

Pipeline leaderboard

Rank ↑	Name	Algorithm	Accuracy (Optimized)	Enhancements	Build time
★ 1	Pipeline 7	Extra Trees Classifier	0.856	HPO-1 FE	00:00:20
2	Pipeline 8	Extra Trees Classifier	0.856	HPO-1 FE HPO-2	00:00:12
3	Pipeline 3	Logistic Regression	0.853	HPO-1 FE	00:00:18
4	Pipeline 4	Logistic Regression	0.853	HPO-1 FE HPO-2	00:00:08
5	Pipeline 2	Logistic Regression	0.838	HPO-1	00:00:02
6	Pipeline 1	Logistic Regression	0.830	*None*	00:00:01
7	Pipeline 6	Extra Trees Classifier	0.823	HPO-1	00:00:05

Experiment summary | **Pipeline comparison**

Rank by: Accuracy (Optimize... ⌄   Score: Cross validation | Holdout

Metric chart ⓘ

Prediction column: output

after run experiment just save model and promote it to deployment space. so that your model deployed and ready to test for experiment.

IBM **Cloud Pak for Data**   Q Search in your workspaces   Buy ⓘ ⓘ   aravind Reddy's Account ⌄   AR

Projects / project_heart / heart project - P8 Snap Random ...

heart project - P8 Snap Random Forest Classifier

Promote to deployment space ⟋

Last modified at Jul 18, 2022 2:50 PM

Description ✎
No description provided.

Created
Jul 16, 2022 11:36 AM

Type
wml-hybrid_0.1

Model ID
c19a9039-774a-458d-9fd9-6... 🗐

Software specification
hybrid_0.1

Hybrid pipeline software specifications
autoai-kb_rt22.1-py3.9

Tags ✎
Add tags to make assets easier to find.

Input Schema

Input

| Column	↑	Type
AGE		"integer"
AVGHEARTBEATSPERMIN		"integer"
BMI		"integer"
CHOLESTEROL		"integer"
EXERCISEMINPERWEEK		"integer"

to test model using Auto AI user interface is far easy. just put values in fields then click on Add to list and then press Predict button and result will show in json format.


This is the prediction done and been tested out in deployed space been provided

Next Deploy the model as a web server and generate scoring End Point. For that we need to open the Node red Application in the Watson studio CatLog list and launch the node red app and install the dashboard from the list.
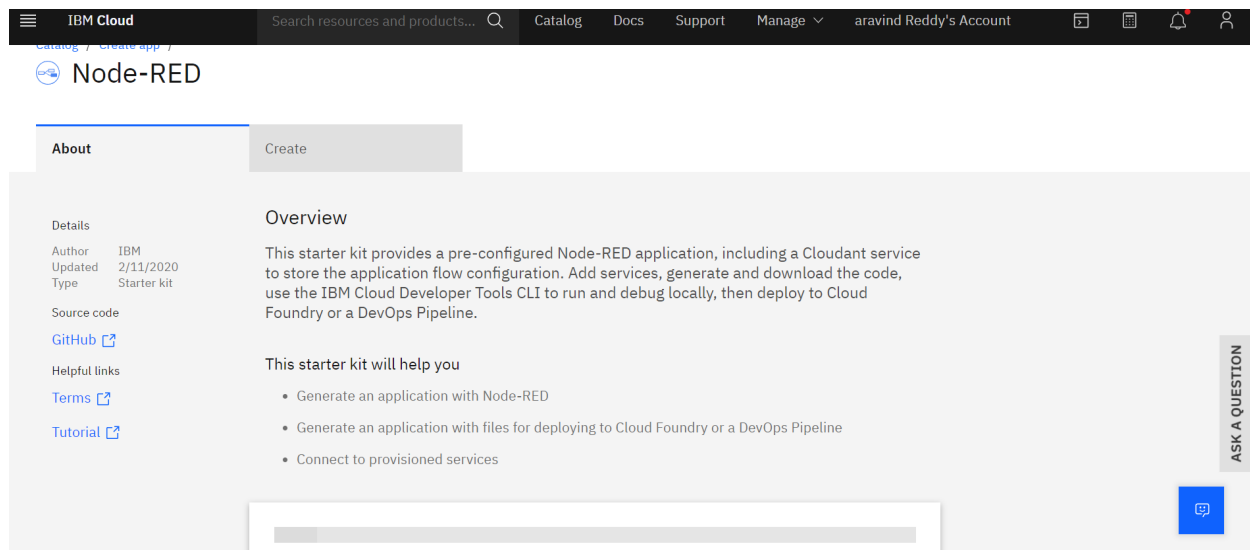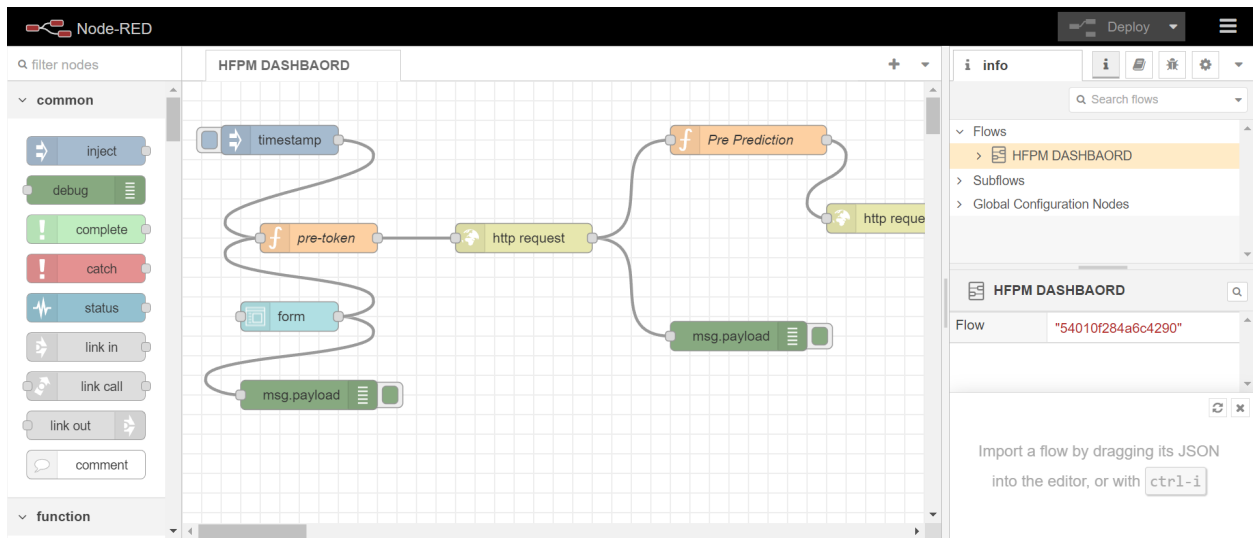
Before that save the API reference and cloud API reference and copy the path over for the reference.

Open the node red application and launch it using and URL provided and use the saved Json file for loading.
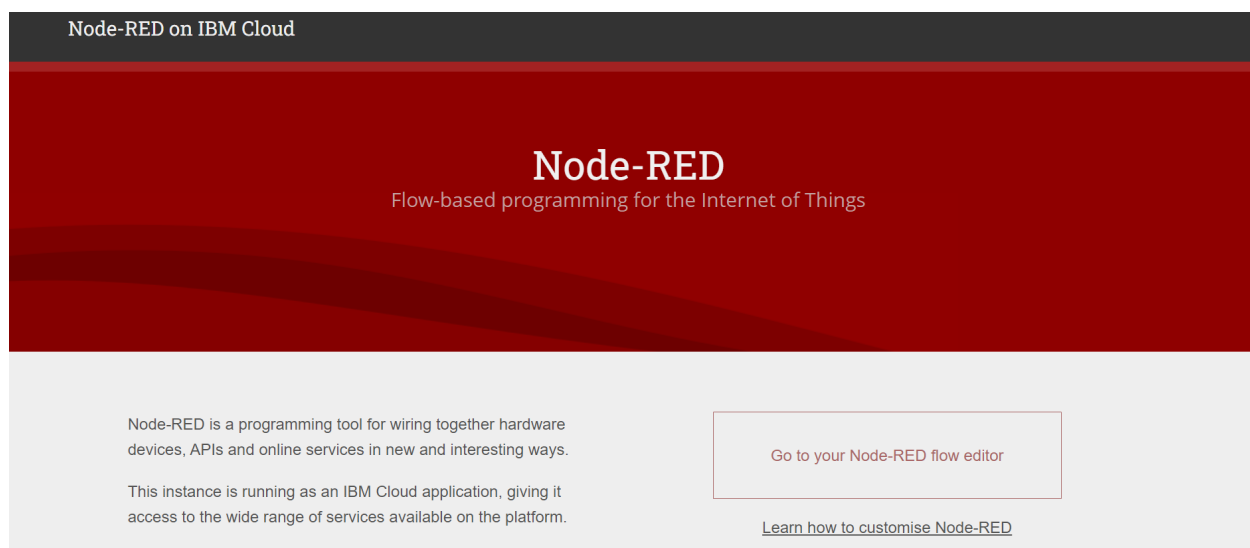
Click on GO to your node red flow editor

Opening it

Deploy the model for the result.

FLOWCHART

This is the project flow to complete the project

1. Log in to IBM account

2. Create IBM Watson Studio and Node-RED Service

3. Create a Watson studio project

4. ADD Auto AI Experiment

5. Run the Auto AI Experiment to build a Machine learning model on the desired dataset

6. Save the model

7. Deploy the model as a web server and generate scoring End Point

8. Create a WEB application Using Node-RED to take user input and showcase Prediction on UI

RESULT:

After loading the json file on and then deploy the model on the Node red application then click on dashboard and then click on the button to launch for the results provide the data and check the prediction

heart rate prediction

AVERAGE HEART BEATS (Per Minute) *

PALPITATIONS PER DAY *

CHOLESTEROL *

BMI *

AGE *

SEX (M or F) *

FAMILY HISTORY (Y or N) *

SMOKER (In Last 5 Years : Y or N) *

EXERCISE (Minutes Per Week) *

SUBMIT	CANCEL

Prediction   **Not at Risk**

Score   **0.9215847778846236**

ADVANTAGES & DISADVANTAGES:

Advantages

User can search for doctor's help at any point of time.

User can talk about their Heart Disease and get instant diagnosis.

Doctors get more clients online.

Very useful in case of emergency.

Disadvantages

Accuracy Issues: A computerized system alone does not ensure accuracy, and the warehouse data is only as good as the data entry that created it.

The system is not fully automated, it needs data from user for full diagnosis.

APPLICATIONS:

Here we use some intelligent data mining techniques to guess the most accurate illness that could be associated with patient's details. Based on result, the can contact doctor accordingly for further treatment. The system allows user to view doctor's details too. The system can be used for free heart disease consulting online.

CONCLUSION:

We can predict the heart disease of any particular age of a person whose data is known for us and get the result in a second that he/she is at a risk or not in percentage also.

FUTURE SCOPE:

The same way we can predict the failure of other body parts of a person based on the food habits and work culture and the area been living.

For that we need to do data mining and collect the several data and apply different data mining techniques to help health care professionals in the diagnosis Based on result, they can contact doctor accordingly for further treatment.

Bibilography:

1.K. Polaraju, D. Durga Prasad, "Prediction of Heart Disease using Multiple Linear Regression Model", International Journal of Engineering Development and Research Development, ISSN:2321-9939, 2017.

2.Marjia Sultana, Afrin Haider, "Heart Disease Prediction using WEKA tool and 10-Fold cross-validation", The Institute of Electrical and Electronics Engineers, March 2017.

3.Dr.S.Seema Shedole, Kumari Deepika, "Predictive analytics to prevent and control chronic disease", https://www.researchgate.net/punlication/316530782, January 2016.

4.Ashok kumar Dwivedi, "Evaluate the performance of different machine learning techniques for prediction of heart disease using ten-fold cross-validation", Springer, 17 September 2016.

5.Megha Shahi, R. Kaur Gurm, "Heart Disease Prediction System using Data Mining

Techniques", Orient J. Computer Science Technology, vol.6 2017, pp.457-466.

6.Mr. Chala Beyene, Prof. Pooja Kamat, "Survey on Prediction and Analysis the Occurrence of Heart Disease Using Data Mining Techniques", International Journal of Pure and Applied Mathematics, 2018.

7.R. Sharmila, S. Chellammal, "A conceptual method to enhance the prediction of heart diseases using the data techniques", International Journal of Computer Science and Engineering, May 2018.

8.Jayami Patel, Prof. Tejal Upadhay, Dr. Samir Patel, "Heart disease Prediction using Machine Learning and Data mining Technique", March 2017.

9.Purushottam, Prof. (Dr.) Kanak Saxena, Richa Sharma, "Efficient Heart Disease Prediction System", 2016, pp.962-969.