Effective Heart Disease Prediction Using IBM Auto AI

Service

Submission for the partial fulfilment of

Project Build-A-Thon - An Exclusive Program for Faculty of India

July-August-2022

Submitted by

Dr. Sandeep K. Raghuwanshi

(sraghuwanshi.cse@gmail.com)

Introduction

Heart disease has been identified as one of the major causes of loss of human lives in recent time. According to one study the number of adults with heart failure are expected to increase by 46% by 2030 in America. It is a critical disease and very hard to detect. However, with advance technology and involvement of Data Science now we have developed which can predict the possibility of a heart disease with better accuracy and efficiency. Following figure tries to explain the heart failure scenario:[1]

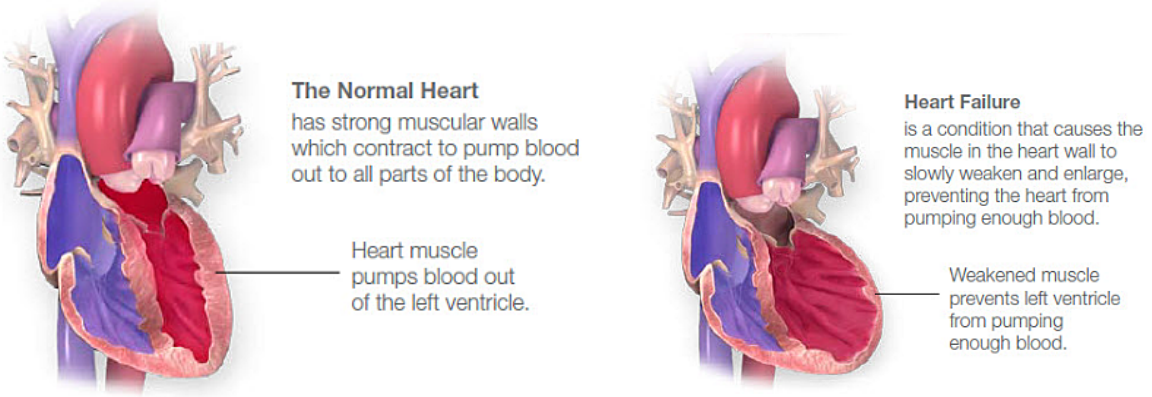

Figure 1 Illustration of a normal heart and a weakened heart that is characteristic of heart failure

Overview of Project

This project is an effort to build an Auto AI based application to predict heart failure of a person based on certain parameters. The project is built using Auto AI on IBM cloud and deployed using IBM Watson Machine Learning services. The Node-Red application was used create a web-based User Interface that allows user to pass inputs and obtain output in the form of prediction label (YES/NO) with score. The following services are used in the development of this project:

- IBM Watson Studio
- IBM Watson Machine Learning

- Node-RED
- IBM Could Storage

Objectives/Purpose

Followings are the key objectives of this project:

a. Work with Watson Studio
b. Create a project in Watson Studio
c. Use Auto Ai experiment to create a model
d. Deploy the ML model as a webserver
e. Integrating Model and Node-RED Service
f. Build an application using Node-RED which takes inputs from the user and showcases the prediction on UI

Literature Survey

Poornima Singh et al. presented a neural network-based model for predicting a risk for heart disease. The proposed model used around 15 medical parameters of a person for the prediction. They tried to understand the relationships between medical factor and heart disease and explored pattern to establish beliefs.[2]

Rohit Bharti et al. uses different machine learning and deep learning techniques to predict a heart failure rate and presented a comparison among them. The dataset used for the study contains 14 parameters and result showed that deep learning achieved around 94.2% accuracy in the prediction of heart failure rate.[3]

Proposed Solution:

Proposed solution is based on IBM Watson machine learning service using AutoAI and Node-RED. Following diagram is presenting the framework of model.

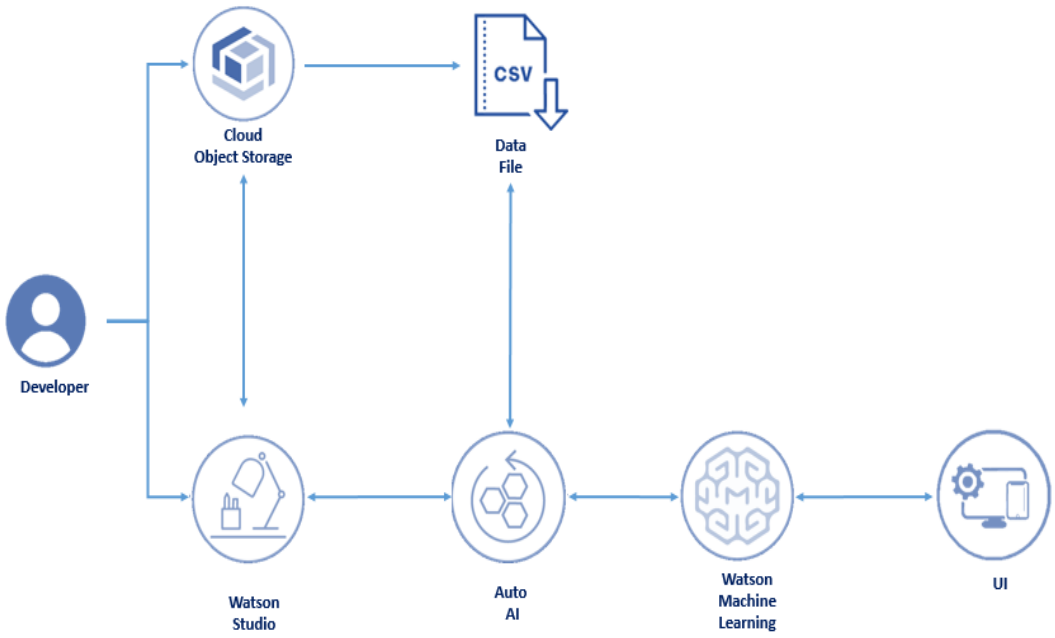

Figure 2 Framework of model

THEORITICAL ANALYSIS

Block diagram Diagrammatic overview of the project

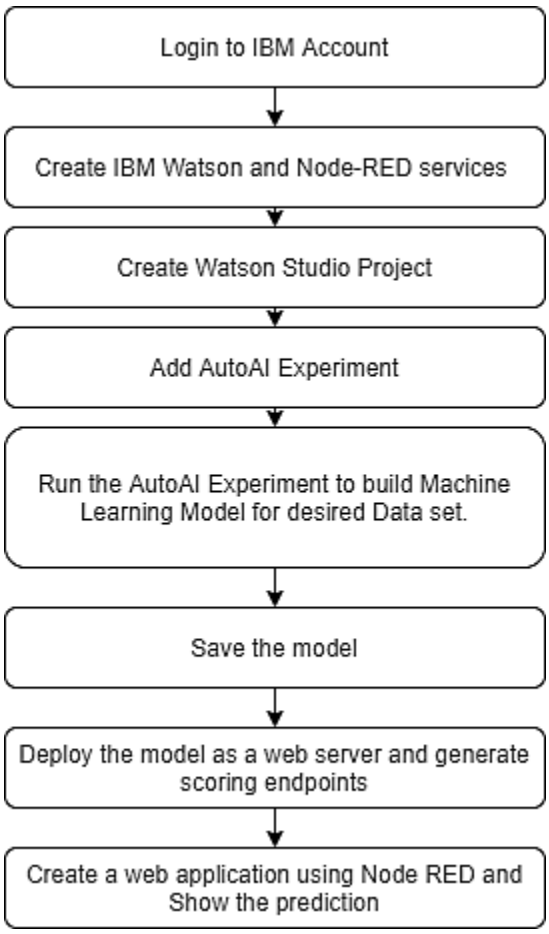

Figure 3 Flow of the Experiment

Hardware / Software Requirements

I have used my laptop machine to compile this project with following specification. However, this project is compiled with IBM cloud and Watson machine learning services. Following the hardware and software specification for the project:

Hardware:

- Processor: Intel(R) Core-i7@ 1.80 GHz.
- RAM -8GB

- 512 GB HDD
- High Speed Internet connection

Software:

- Operating System (Any)
- IBM Watson Studio
- IBM Watson Machine Learning
- Node-RED
- IBM Could Storage

EXPERIMENTAL INVESTIGATIONS

Dataset- The data set is very rich in terms of number of entries available to train the model. The data set contains 10800 entries over ten features. The outcome is HEARTFAILURE, and rest of the parameters are considered as input variables. The summarized view of the dataset is present on following figure.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10800 entries, 0 to 10799
Data columns (total 10 columns):
 #   Column               Non-Null Count  Dtype
---  ------               --------------  -----
 0   AVGHEARTBEATSPERMIN  10800 non-null  int64
 1   PALPITATIONSPERDAY   10800 non-null  int64
 2   CHOLESTEROL          10800 non-null  int64
 3   BMI                  10800 non-null  int64
 4   HEARTFAILURE         10800 non-null  object
 5   AGE                  10800 non-null  int64
 6   SEX                  10800 non-null  object
 7   FAMILYHISTORY        10800 non-null  object
 8   SMOKERLAST5YRS       10800 non-null  object
 9   EXERCISEMINPERWEEK   10800 non-null  int64
dtypes: int64(6), object(4)
memory usage: 843.9+ KB
```

Figure 4 Heart Disease Data set

Experiment was conducted using IBM AutoAI service with Watson Machine Learning. The study was carried over following machine learning algorithms:

- Decision Tree Classifier.
- Extra Tree Classifier.
- Gradient Boosting Classifier.
- LGBM Classifier.

- Logistics Regression.
- Random Forest Classifier.
- Snap Logistics Regression
- Snap Random Forest Classifier
- Snap SVM Classifier
- XGB Classifier

Following diagram presents the training optimization through different pipelines and cross validation score:

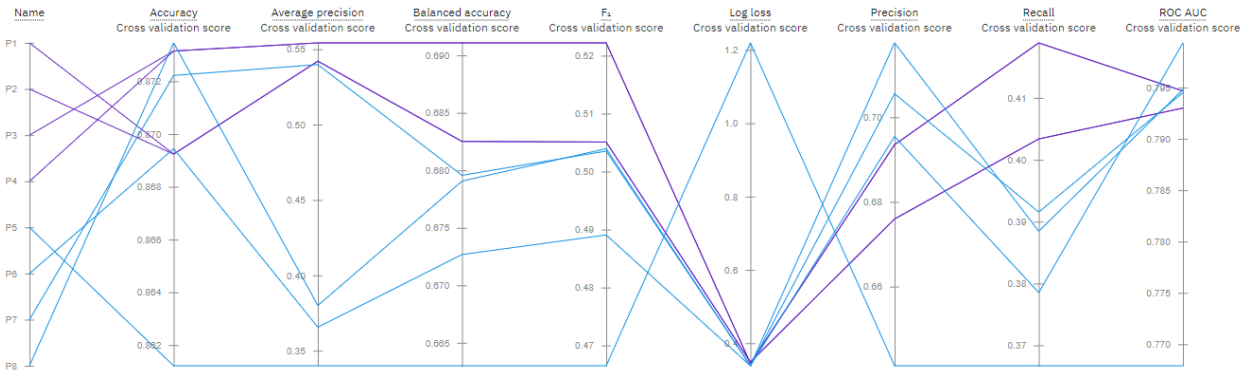

Figure 5 Metric Chart

FLOWCHART

The following figure shows the progression map for the training process of AutoAI service.

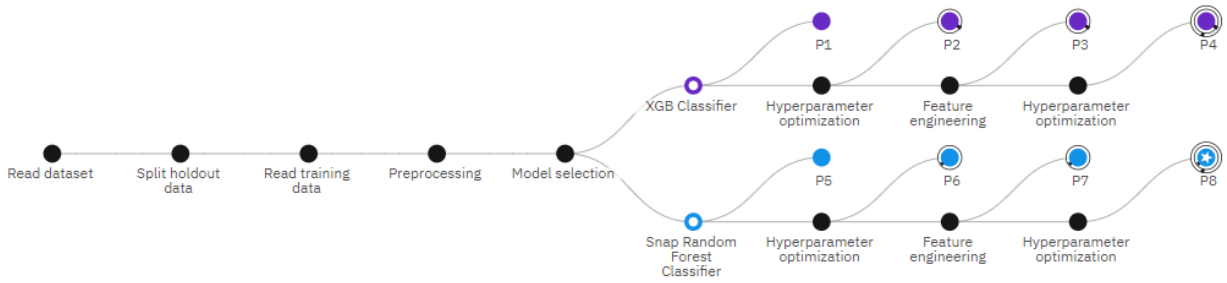

Figure 6 Progress Map for the train Auto AI

RESULT

Figure presents the outcome of the experiment and training process of model.

★	1	**Pipeline 8**	Snap Random Forest Classifier	**0.873**	HPO-1 FE HPO-2	00:01:30
	2	**Pipeline 3**	XGB Classifier	**0.873**	HPO-1 FE	00:00:51
	3	**Pipeline 4**	XGB Classifier	**0.873**	HPO-1 FE HPO-2	00:01:48
	4	**Pipeline 7**	Snap Random Forest Classifier	**0.872**	HPO-1 FE	00:01:04
	5	**Pipeline 6**	Snap Random Forest Classifier	**0.869**	HPO-1	00:00:09
	6	**Pipeline 1**	XGB Classifier	**0.869**	*None*	00:00:01
	7	**Pipeline 2**	XGB Classifier	**0.869**	HPO-1	00:00:14
	8	**Pipeline 5**	Snap Random Forest Classifier	**0.861**	*None*	00:00:01

Figure 7 Outcome of Experiment

From the above result it is clear that the best performing machine learning model is **“Snap Random Forest Classifier”** with 87.3 % accuracy of prediction.

Sample output in GUI form:

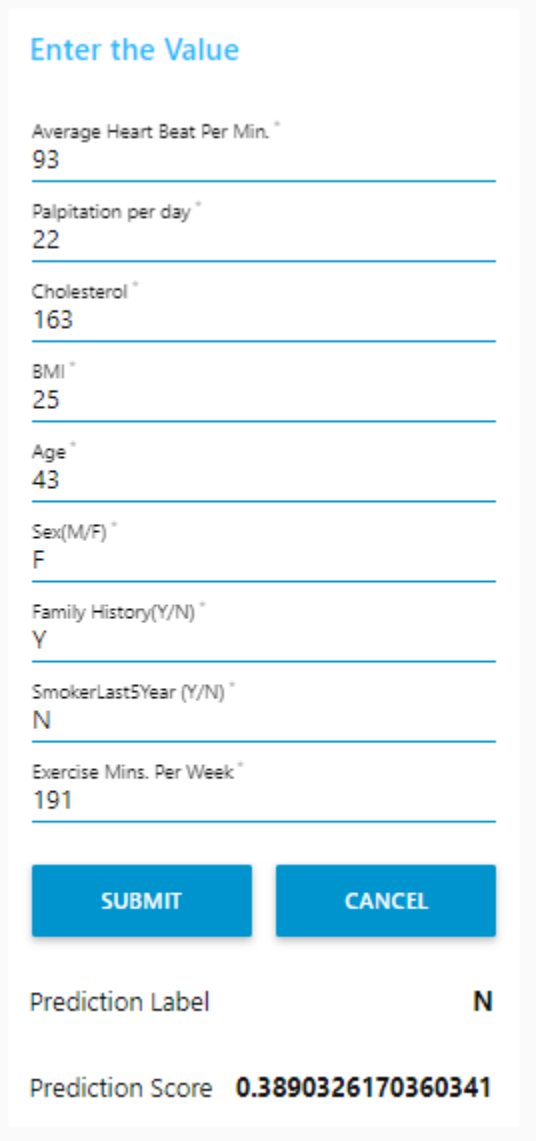

Figure 8 GUI Based output

Conclusion and Future Scope

The final outcome of the experiment is the trained model based on random forest classifier with 87.3% prediction accuracy. The experiment conducted a comparative analysis among various machine learning models and the select the best performing over evaluation metrics.

The future scope can include applying some feature engineering and data imbalancing method to improve model performance.

References

[1] https://www.ibm.com/blogs/research/2017/04/using-ai-to-predict-heart-failure

[2] Singh P, Singh S, Pandi-Jain GS. Effective heart disease prediction system using data mining techniques. Int J Nanomedicine. 2018 Mar 15;13(T-NANO 2014 Abstracts):121-124. doi: 10.2147/IJN.S124998. PMID: 29593409; PMCID: PMC5863635.

[3] Bharti, Rohit, Khamparia, Aditya, Shabaz, Mohammad, Dhiman, Gaurav,Pande, Sagar, Singh, Parneet, “Prediction of Heart Disease Using a Combination of Machine Learning and Deep Learning” Computational Intelligence and Neuroscience, Hindawi, https://doi.org/10.1155/2021/8387680 , 2021