Effective Heart Disease Prediction Using IBM Auto AI Service

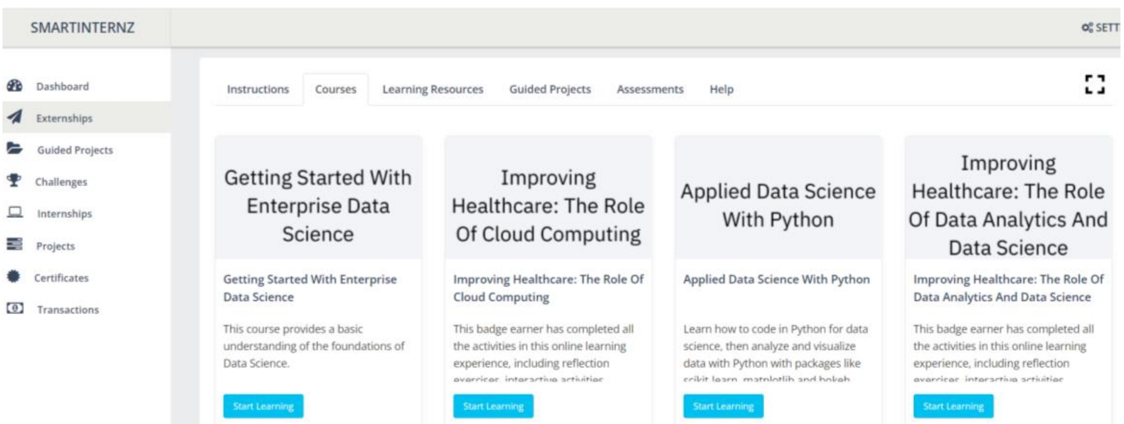
Name : Poonam M. PangarkarA

Email ID: pmpangarkar@apsit.edu.in

A. P. Shah Institute of technology

Note:

1. **I have completed all the following courses but I have received only 2 certificates from credly.com**



2. **I have completed my project using**

User ID

pmpangarkar@apsit.edu.in

Password

****

Email

pmpangarkar@apsit.edu.in

Effective Heart Disease Prediction Using IBM Auto AI Service

1 INTRODUCTION

1.1 Overview

Heart failure is a common event caused by Cardiovascular diseases which causes major death count and several diagnosis methods were also involved. But still the failure rate prediction is lacking because of medical examination as well as tools used. This paper explores the meticulousness of a machine learning and artificial intelligence based automatic prediction model, which is built by IBM services for heart failure rate prediction where the dataset is trained and a model is built. The auto AI instance is created in the IBM Watson Studio and machine learning services are linked with it. The auto AI service determines the best algorithm as the Gradient Boost algorithm for the given dataset here and automatically classifies it as a binary classification problem with values as Y/N for heart failure. Several algorithms can be chosen and deployed. The NodeRED service is used to deploy the model as a final application. The accuracy along with precision and recall measures and metrics were chosen automatically by the system as best ones. The infographics of the results determines that several other algorithms can also be merged and executed one. Also it is evident from the results, that with a minimum span of time, the application is automatically modeled and deployed for the major threatening disease.

1.2 Purpose

T Human heart is an important organ in the human body that is responsible for blood circulation ie, for oxygen and energy supply to all organs of the body. Heart disease leads to abnormal blood circulation in the body that might be fatal for human life. Heart disease is the main cause of death according to the Centers for Disease Control (CDC) in the entire universe. There are two categories of risk factors for heart diseases. Alterable risk factors, e.g., cigarette usage and workouts, and non-alterable risk factors, e.g., gender, age factor and generation history. By using conventional medical methods, it is much difficult to determine the symptoms of heart failure and also complex, costly and time consuming [5]. Angiography is one of the best method of medical tests for the diagnosis of heart failure. But it has other effects as well as costlier and demands extraordinary technical expertise. However, accuracy of prediction is a sustaining problem in these systems. Therefore, to overcome the angiography issues, data mining and ML techniques are much helpful. There are also systems which monitors the parameters like blood pressure, ECG, SpO2 and also relevant environmental indicators.

2 LITERATURE SURVEY

2.1 Existing problem

A Machine learning technique was developed by Dr.M.Thyagaraj et al [1] where the preliminary process is to gather heart data base and then to reduce data redundancy and improve data integrity, the data normalization is performed by using Zero-Score (Z-Score). Then, attribute reduction is performed by using soft computing techniques namely Genetic Algorithm, Particle Swarm Optimization, Crow Search Optimization and

Opposition Based Crow Search Optimization. Finally, the Radial Basis Function-Transductive Support Vector Machines (RBF-TSVM) classifier is used for heart disease prediction. It is found that the proposed OCSO technique attains superior performance related with the existing method in terms of accuracy, specificity and sensitivity. The datasets is Cleveland Heart Disease Dataset (CHDD) accessible on the UCI Repository with 303 patient records in the database.

Mr.Sumit et al [2] developed a model named Optimized DNN using Talos which was found to produce a high accuracy compared to other models. Steps which they followed was Data cleaning to filter and modify the data, so that it is easier to understand. Feature engineering to extract some features data set which provides accuracy and is a process of retrieving features data from raw data to improve the quality of the model. Multiple learning algorithms were used and they devised the Hyper-parameter optimization- Talos algorithm which follows the steps of (Prepare, Optimize, Deploy) POD. Prepare (P) - hyper parameter space for the experiments to choose the optimization strategy are defined. Optimize (O) - determining an optimal hyper parameter combination for a given prediction task. Deploy (D)- required assets are sorted locally for production purpose from local and remote deployment of a model. Report generation and evaluating is the final step which has multiple options for analysis and evaluation of experiments including all plots of visual analysis for experimental progress. All the methods were compared and deep learning neural networks (DNN) with respect to Talos optimization isfound to be accurate by 90.76%.

Another method was proposed by Ashir et al [3] where a feature selection method for better heart risk prediction accuracy is implemented, that uses an adaptive size floating window for feature elimination (FWAFE). Gridsearch algorithm is used to find the neural network's optimal configuration. It was found that DNN trained with new learning algorithms with multiple hidden layers shows better performance. Cleveland heart disease database with 303 features is used here. FWAFE-ANN and FWAFE-DNN -2 hybrid systems are implemented i.e using artificial neural network (ANN) and deep neural network (DNN). The optimal subset of features are developed every time. The grid search algorithm is used for optimization. Next, a feature subset is built using FWAFE, whereas for classification purpose DNN is used. Multiple classifier algorithms are used and experimental results confirm that the proposed model outperformed several other machine learning techniques for heart disease diagnosis with accuracy range of 50.00–91.83%.

Mobile cloud computing approach is proposed by Venkatesan et al [7] which is superior to health monitoring methods because of the centralized cloud data access and reporting to mobile phones. In their method, the risk assessment of ECG tele monitoring and coronary heart disease (CHD) risk are merged using mobile cloud computing. Mobiles receive ECG data from different sensing devices without any processing. In feature extraction process, the heart rate variability (HRV) of the ECG signal is determined by analyzing R-peaks using wavelet transform. Since the cloud approach in mobiles deals with huge volume of data, MIT–BIH arrhythmia database with 160 files has been used. The performance of the resultant classifier is evaluated and compared with other similar classifiers and found that proposed methodology - ANFIS classifier generates a good accuracy of 98.75%.

Effective Heart Disease Prediction Using IBM Auto AI Service

In this work, an automatic predictor system forpredicting heart failure is built using IBM auto AI service. Itbuilts a classification model automatically based on inputs and classifies the model using the gradient boostingclassifier with an accurate result. It classifies with an accuracy value of 0.874 by involving boosting measuresalso. The model is deployed using NodeRED service and theresults are visualized in the output.

2.2 Proposed solution

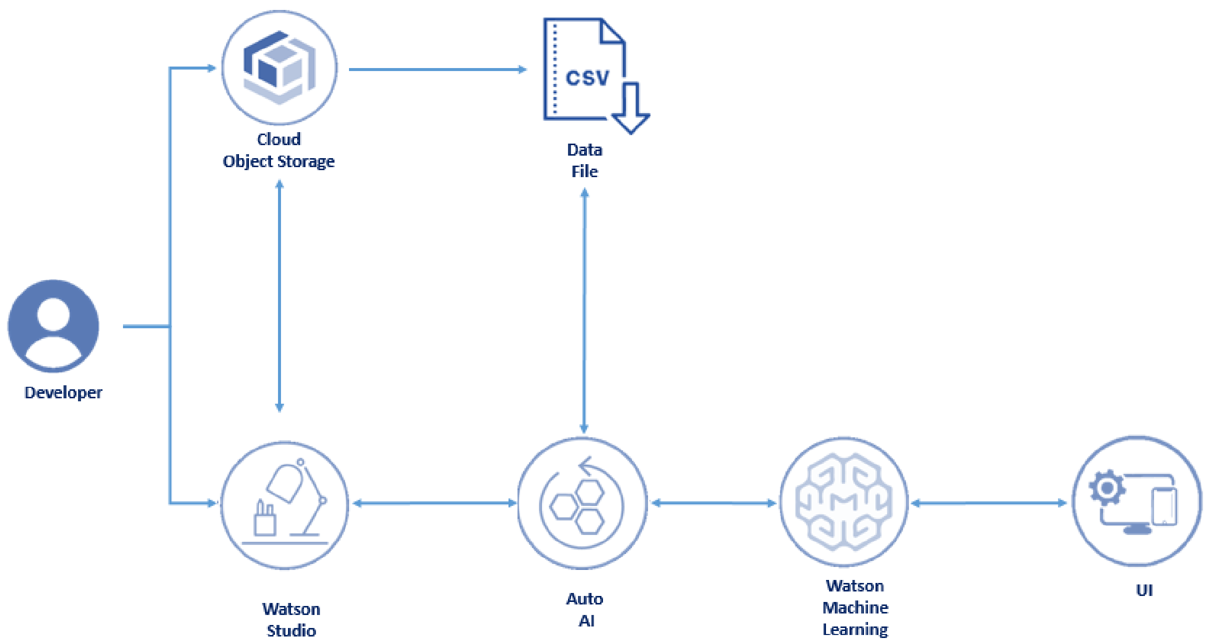
The main focus of this system is to automate the prediction of heart failure by using the AutoAI service where separate data preprocessing, feature extraction steps and development need not be processed. Auto AI automates Data preparation, Model development, Feature engineering and Hyper-parameter optimization. The service is available in IBM Watson Studio[8]. This makes data scientists to quickly get started and expert data scientists to speed up the experimentation time drastically. It provides a multimodal data science and AI environment where data and analytics specialists collaborate with other experts and optimize model performance. The benefits of auto AI are to select the top performing models and algorithms faster and also to start up the execution quickly. It also maintains consistency and integrity of end-end AI and ML environment. The key feature of it is the integrated User Interface where data preparation is automated and optimization of hyper parameters are done. The deployment is also much simpler, where the execution is happening in just few clicks of services. The entire AI or ML lifecycle is automated which is really a biggest advantage. As AutoAI analyzes the dataset and discovers data transformations, algorithms, and parameter settings these model pipelines are created iteratively. The architecture of the proposed IBM AutoAI is shown in Fig.
The overall process flow is depicted in Fig. 2. The Services Used for the heart failure detection are IBM Watson Studio, IBM Watson Machine Learning, Node-REDand IBM Cloud Object Storage. The dataset contains 9 features that can be used to predict mortality by heart failure. A model is built using IBM Auto AI service and a web application is built where we can get the prediction of heart failure. An AutoAI experiment is created in Watson studio and AutoAI experiment in the Asset page in software option is proceeded. The model is a machine learning model and need algorithms, measures and metrics dealing with that. So, it is associated with the machine learning instance.

The dataset used is a repository from IBM used for the purpose of heart failure prediction. It has 10 columns with 10800 rows with various features causing heart failure like smoking history, age, palpitation count, gender, walking habit etc., With the dataset, the model is to be trained where AutoAI analyses it and classifies it automatically as Y/N based on the inputs. Among multiple algorithms built in the system, it automatically trains the model and producesan output as the given problem as a Binary Classification problem. The performance measures are also inbuilt and the best measure for the model chosen by auto AI system is accuracy. By default, top two best classification algorithms for the model are executed but it is customizable. So, the settings are customized for running top four algorithms and it is executed. On execution, it can be seen that four pipelines for each algorithm is executed and the map can also be visualized.

Effective Heart Disease Prediction Using IBM Auto AI Service

3 THEORITICAL ANALYSIS

3.1 Block diagram



3.2 Hardware / Software designing

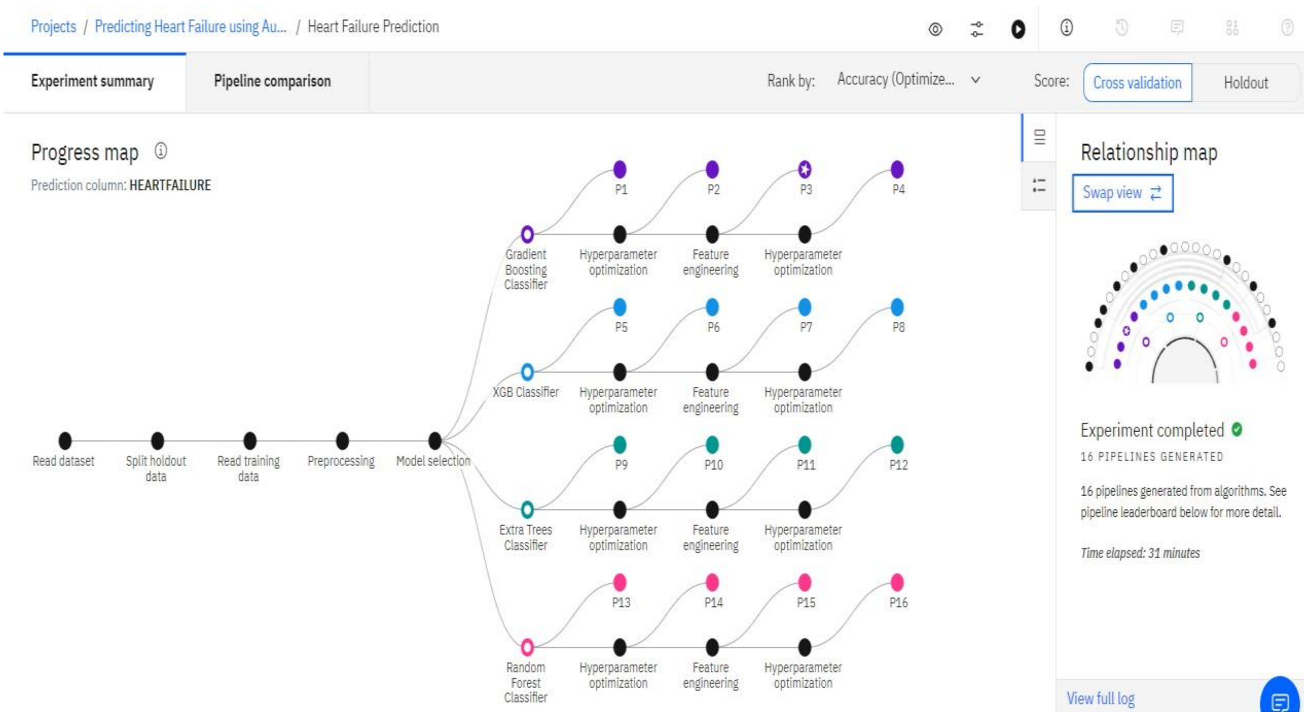
Hardware Requirements

1. Processor:Intel core i3 processor
2. RAM: 4GB
3. System Type: 64-bit operating system

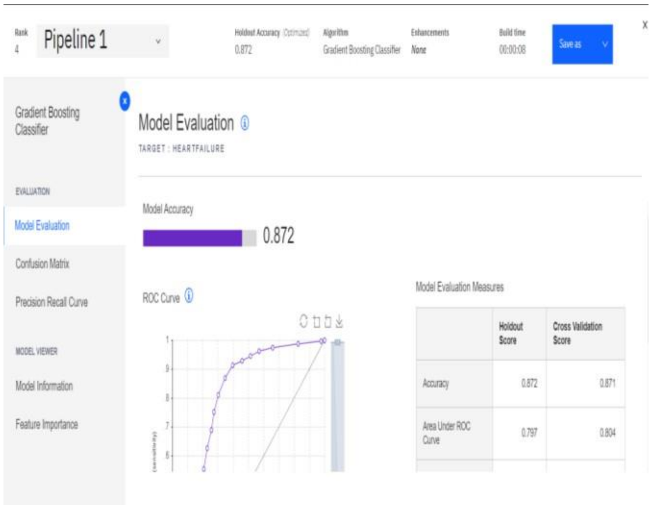
Software requirements:

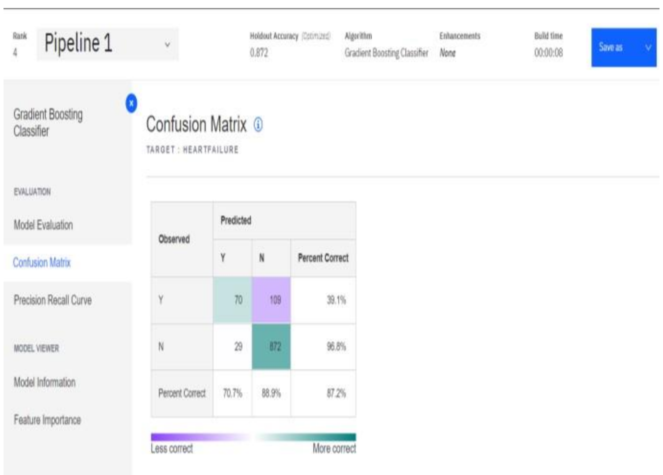
1. IBM Waston Studio
2. IBM Waston Machine Learning
3. Cognos
4. Node-RED
5. IBM Cloud Object Storage

4 EXPERIMENTAL INVESTIGATIONS

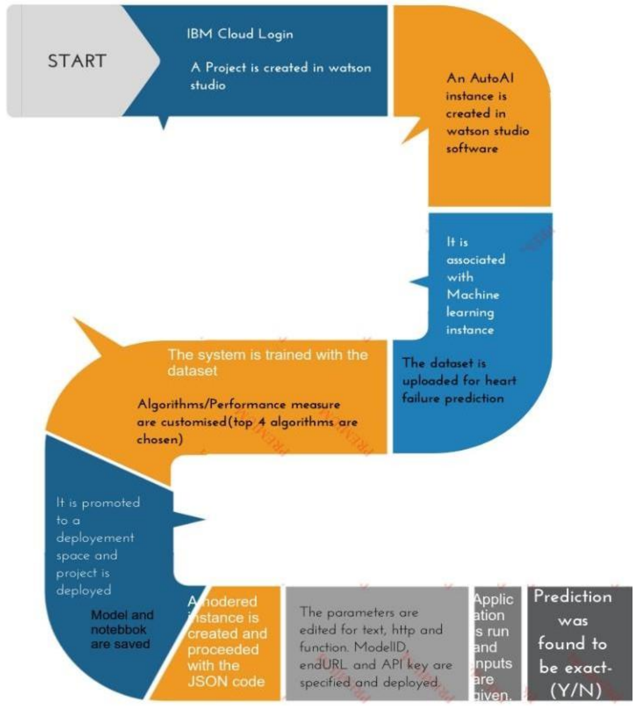


Rank	Pipeline	Algorithm	Enhancement Algorithm Used	Accuracy
1	Pipeline 3	Gradient BoostingClassifier	HPO-1 andFE	0.874
2	Pipeline 4	Gradient BoostingClassifier	HPO-1, FE and HPO-2	0.874
3	Pipeline 2	Gradient BoostingClassifier	HPO-1	0.872
4	Pipeline 1	Gradient BoostingClassifier	None	0.872
5	Pipeline 8	XGB Classifier	HPO-1, FE and HPO-2	0.865
6	Pipeline 7	XGB Classifier	HPO-1 andFE	0.871
7	Pipeline 5	XGB Classifier	None	0.865
8	Pipeline 6	XGB Classifier	HPO-1	0.863
9	Pipeline 9	Extra Trees Classifier	None	0.856
10	Pipeline 10	Extra Trees Classifier	HPO-1	0.856

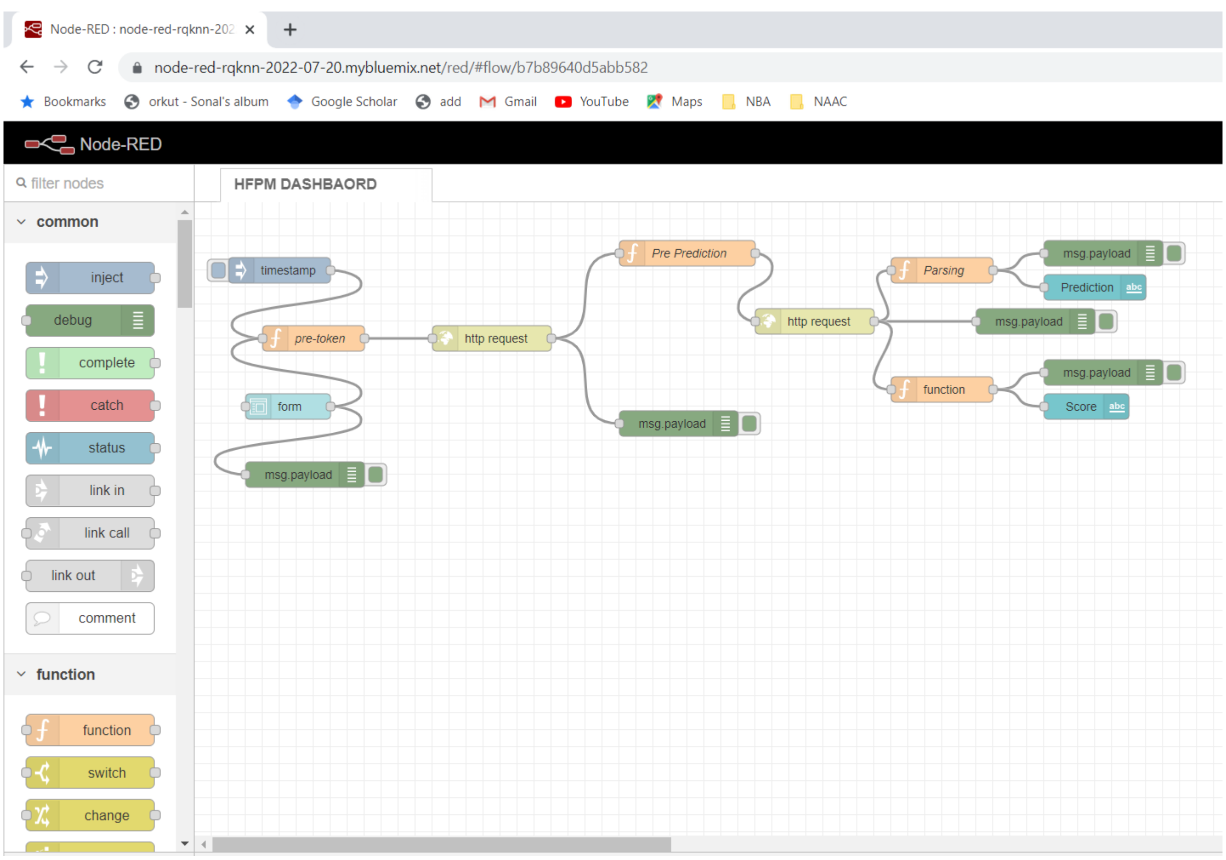

Pipeline 1
Holdout Accuracy (Optimized)
0.872
Algorithm
Gradient Boosting Classifier
Enhancements
None
Build time
00:00:08
Save as
Gradient Boosting Classifier
EVALUATION
Model Evaluation
Confusion Matrix
Precision Recall Curve
MODEL VIEWER
Model Information
Feature Importance
Model Evaluation
TARGET : HEARTFAILURE
Model Accuracy
0.872
ROC Curve
Model Evaluation Measures
Holdout Score
Cross Validation Score
Accuracy
0.872
0.871
Area Under ROC Curve
0.797
0.804


Pipeline 1
Holdout Accuracy (Optimized)
0.872
Algorithm
Gradient Boosting Classifier
Enhancements
None
Build time
00:00:08
Save as
Gradient Boosting Classifier
EVALUATION
Model Evaluation
Confusion Matrix
Precision Recall Curve
MODEL VIEWER
Model Information
Feature Importance
Confusion Matrix
TARGET : HEARTFAILURE
Observed
Predicted
Y
N
Percent Correct
Y
70
109
39.1%
N
29
872
96.8%
Percent Correct
70.7%
88.9%
87.2%
Less correct
More correct

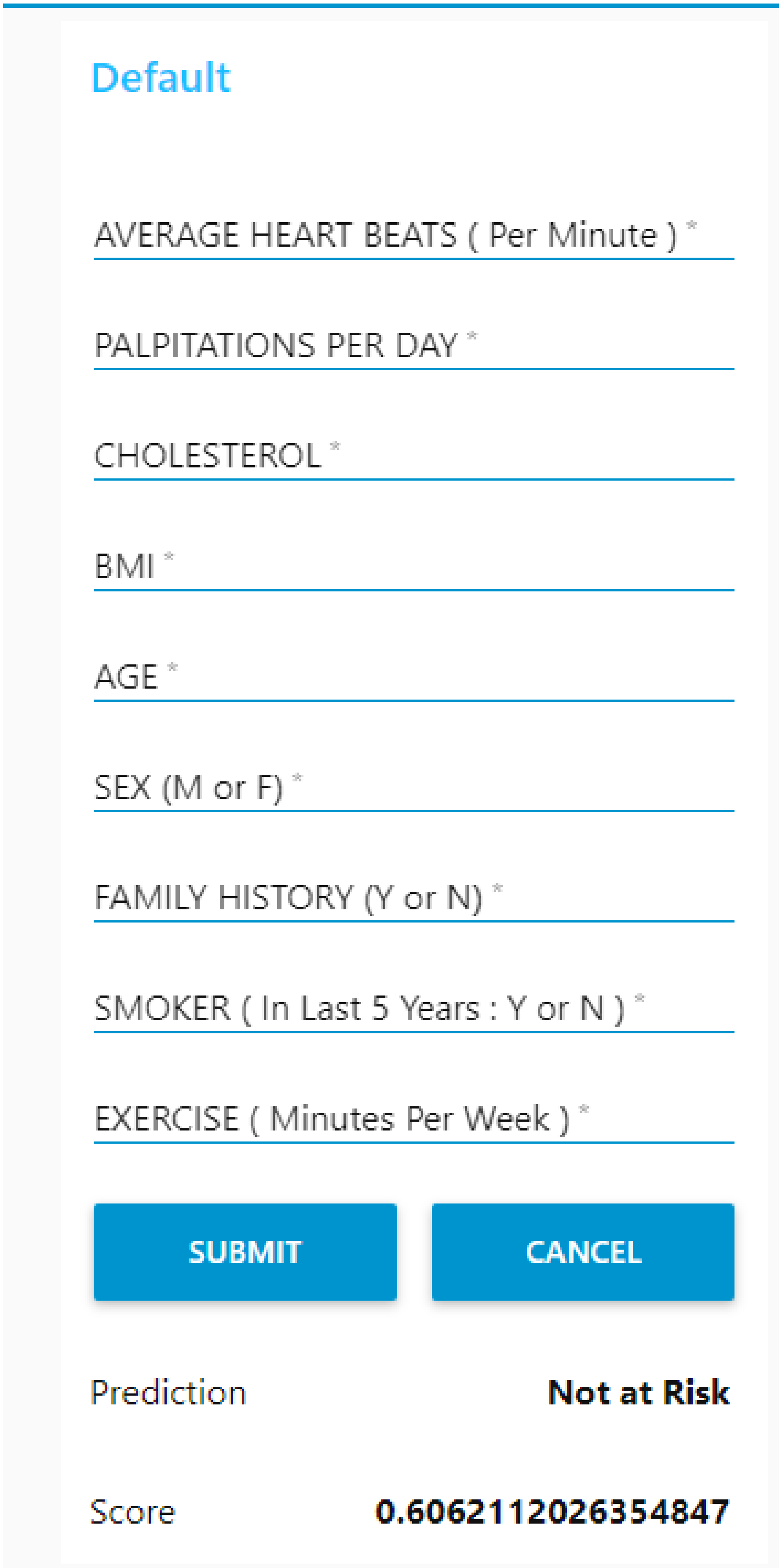
5 FLOWCHART


START
IBM Cloud Login
A Project is created in watson studio
An AutoAI instance is created in watson studio software
It is associated with Machine learning instance
The dataset is uploaded for heart failure prediction
The system is trained with the dataset
Algorithms/Performance measure are customised(top 4 algorithms are chosen)
It is promoted to a deployement space and project is deployed
Model and notebbok are saved
A nodered instance is created and proceeded with the JSON code
The parameters are edited for text, http and function. ModelID, endURL and API key are specified and deployed.
Application is run and inputs are given.
Prediction was found to be exact-(Y/N)

6 RESULT


Node-RED : node-red-rqknn-202
node-red-rqknn-2022-07-20.mybluemix.net/red/#flow/b7b89640d5abb582
Bookmarks
orkut - Sonal's album
Google Scholar
add
Gmail
YouTube
Maps
NBA
NAAC
Node-RED
filter nodes
HFPM DASHBAORD
common
inject
debug
complete
catch
status
link in
link call
link out
comment
function
function
switch
change
timestamp
pre-token
http request
form
msg.payload
Pre Prediction
http request
msg.payload
Parsing
msg.payload
Prediction
msg.payload
function
msg.payload
Score

Effective Heart Disease Prediction Using IBM Auto AI Service



Default

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 **0.6062112026354847**

7 ADVANTAGES AND DISADVANTAGES

7.1 Advantages

1. Proposed model emerges as the best ML image classification technique on account of having high accuracy.
2. Comparing with other algorithms, it requires less amount of image pre-processing.
3. It reduces images without losing features making them much easier to process by using relevant filters and reusability of weights resulting in good prediction
4. It can learn to perform any given task automatically by learning from the training data i.e. prior knowledge is not necessary.

7.2 Disadvantages

1. Proposed model requires a huge set of training data to get better results.
2. An appropriate model is required.
3. Total process is time consuming
4. Process is more tedious and exhaustive procedure.

8 CONCLUSION

The prediction of heart failure is one of the mostly needed requirement in the medical field. Here, without any coding, automatically built IBM services are utilised and a Binary classification AutoAI model is executed successfully. Four algorithms were chosen and so 16 pipelines were executed. The best performance measure which the model resulted is accuracy with the Gradient Boosting Algorithm with a best result of 0.874. The build time 00:00:33 seconds, which is saved as model and deployed. The model is further tested and integrated with Node Red service for interfacing and creating a web service. It is seen that the model exactly predicted the class Y/N representing the possibility of Heart Failure/No Heart Failure. Thus, the IBM Watson studio, AutoAI, NodeRED and Cloud services are utilised and a model is automatically deployed successfully without any implementation of coding.

This can also be used to create an awareness for heart failures. So instead of detecting failures, an application can be created to be a preventive measure for failures. Several other services in IBM can also be utilised where a real time mobile application can be developed. The Chat bots, Natural Language disambiguations, Sensor based cloud applications can also be developed which is automatic and fast.

9 FUTURE SCOPE

1. Test accuracy can be improved

2. Computation time can be decreased with the help of classifier boosting techniques such as fine-tuning hyperparameters, using an even larger dataset with added data augmentation, adding several appropriate layers , training longer by using more epochs, etc.

3. The use of classifier boosting can make the model faster in which the first model is built from the training dataset whereas a second model is built correcting the errors of the first model.

4. Thus all these techniques can help raise the accuracy of the model and the tool will be significant asset in various medical facilities operating on heart disease.

10 BIBILOGRAPHY

[1] Mr.M. Thyagaraj & Dr.G. Suseendran, Vels Institute of Science, Technology & Advanced Studies (VISTAS), Chennai, " A Proficient System to Identify Heart Diseases with the Aid of Artificial Intelligence and Soft Computing Techniques", Journal of Adv Research in Dynamical & Control Systems, Vol. 11, 06-Special Issue,2019

[2] Sumit Sharma, Mahesh Parmar, "Heart Diseases Prediction using Deep Learning Neural Network Model", International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-9 Issue-3, January 2020

[3] Ashir Javeed,Sanam Shahla Rizvi, Shijie Zhou, Rabia Riaz, Shafqat Ullah Khan and Se Jin Kwon , "Heart Risk Failure Prediction Using a Novel Feature Selection Method for Feature Refinement and Neural Network for Classification", Hindawi Mobile Information Systems Volume 2020, Article ID 8843115,https://doi.org/10.1155/2020/8843115

[4] Khaled Mohamad Almustafa, "Prediction of heart disease and classifiers sensitivity analysis" - BMC Bioinformatics (2020) 21:278 https://doi.org/10.1186/s12859-020-03626-y

[5] Aamir Shahzad, Yang Sun Lee, Malrey Lee, Young-Gab Kim and Naixue Xiong, "Real-Time Cloud-Based Health Tracking and Monitoring System in Designed Boundary for Cardiology Patients", May 2018, Hindawi Journal of Sensors Volume 2018, Article ID 3202787, 15 pages https://doi.org/10.1155/2018/3202787

[6] Chao Lia, Xiangpei Hua, Lili Zhangb, "The IoT-based heart disease monitoring system for pervasive healthcare service", International Conference on Knowledge Based and Intelligent Information and Engineering Systems, KES2017, September 2017, ScienceDirect Available online at www.sciencedirect.com, Procedia Computer Science 112 (2017) 2328–2334, Elsevier

[7] C. Venkatesana, P. Karthigaikumar, S. Satheeskumaran, "Mobile cloud computing for ECG

telemonitoring and real-time coronary heartdisease risk detection", April 2018, Biomedical Signal Processing andControl, 44-2018 – 138-145

[8] ibm.cloud.com

[9] Dataset source : https://github.com/IBM/predictive-model-on-watson-ml/blob/master/data/patientdataV6.csv