Deployment of Heart Disease Prediction Model in Cloud Environment

I. BHUVANESHWARRI

Assistant Professor (Senior), Department of Information Technology, Institute of Road and Transport Technology, Erode, Tamilnadu, India

Abstract- ***The main reason for heart failure is Cardiovascular Diseases (CVDs). The dataset used in this paper contains 9 attributes that can be used to predict death or mortality by heart failure. In this paper, a prediction model in cloud environment is built to display the prediction outcome of the heart failure. The cloud service automatically generated the effective heart disease prediction model using pipeline based approach. In this proposed work, Snap Random Forest Classifier is selected as the effective heart disease prediction model among other 7 prediction models with classification accuracy of 87.3%. The primary objective of this effective heart disease prediction model is to determine whether a patient should be diagnosed with heart disease or not, which is a binary outcome either 0 or 1. The outcome of binary value 1 implies that the patient will be diagnosed with heart disease and outcome of binary value 0 implies that the patient will not be diagnosed with heart disease.***

I. INTRODUCTION

Heart related diseases or Cardiovascular Diseases (CVDs) are the main reason for a huge number of death in the world over the last few decades and has emerged as the most life- threatening disease, not only in India but in the whole world. So, there is a need of reliable, accurate and feasible system to diagnose such diseases in time for proper treatment. Machine Learning algorithms and techniques have been applied to various medical datasets to automatethe analysis of large and complex data. Many researchers, in recent times, have been using several machine learning techniques to help the health care industry and the professionals in the diagnosis of heart related diseases.

II. RELATED WORKS

It is very important to take into account the prediction of risk level of heart disease for healthcare industry in order to ease the medical treatment for the patients. Data science classification techniques are used in a number of applications like healthcare analytics, customer analytics, marketing analytics, water quality analytics, textile production analytics,manufacturing analytics and textile waste analytics etc., There are various heart disease prediction models are available based on data mining techniques such as regression, clustering, association rule and classification techniques such as decision tree, naïve Bayes, random forest, artificial neural network etc., Even though there are lot of prediction models and ensemble techniques available, there is no single infrastructure or framework existing to execute all the above techniques altogether. Therefore, for developing the effective and best heart disease prediction model, lot of efforts are needed to incorporate everything. Ramalingam et al.,[1] provided the survey about the heart disease prediction using machine learning techniques. They discussed about algorithms, techniques and performance of various models such as Support Vector Machines (SVM), K-Nearest Neighbour (KNN), NaïveBayes, Decision Trees (DT), Random Forest (RF) and ensemble models. Mohan et al., [2] also proposed effective heart disease prediction using hybrid machine learning techniques. Their method aimed at finding significant features by applying machine learning techniques resulting in improving the accuracy in the prediction of cardiovascular disease. Their prediction model is introduced with different combinations of features and several known classification techniques. They produced an enhanced performance level with an accuracy level of 88.7% through the prediction model for heart disease with the hybrid random forest with a linear model (HRFLM). Various research workers

like Patel et al.,[3], Jagtap et al.,[4],Khourdifi& Bahaj [5] and Gavhane [6] have provided different solutions using different machine learning techniques for developing effective heart disease prediction model.

III. METHODS

3.1 Proposed Solution

The proposed solution is to develop a effective prediction model using IBM AutoAI service. Under AutoAI, there are various machine learning techniques are available. Using pipeline concept, various prediction models with different machine learning techniques are developed and also found effective prediction model among them. Then finally, with the effective prediction model, a web server application is also built to showcase the outcome of prediction of heart failure using node RED service.

3.2 Architectural flow diagram

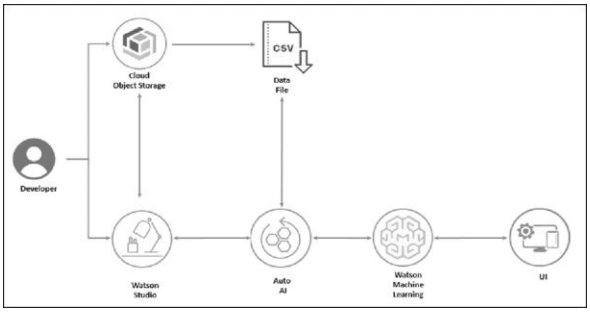


Figure.1 Architectural flow diagram of Effective Heart Disease Prediction Model

Figure.1 depicts the architectural flow diagram of the proposed effective heart disease prediction model using IBM Auto AI service. The IBM cloud provides SaaS (Software-As-A- Service).Using this service, the developer creates the IBM cloud account to acquire the services provided by the IBM cloud. Initially, the developer creates cloud object storage service and Watson studio services. After that create Node RED service for deploying web server application. Then create Watson machine learning service and create Auto AI experiment with thepatient dataset to build a machine learning model. Now the model is ready to deploy as webserver and generate scoring end point. Then using Node RED service, create User Interface for accessing the web server application by the user and create web application to take user input and display prediction result on User Interface.

IV. EXPERIMENTAL RESULTS

4.1 EXPERIMENTAL

In this paper, the dataset named patientdataV6.csv is used. It can be downloaded from github website. This dataset contains 10 attributes and 10,800 patient's sample records. Out of 10 attributes, 9 attributes are conditional /independent attributes and 1 is decision/dependent attribute. The conditional attributes are AVG HEARTBEATS PER MIN, FAMILY HISTORY, PALPITATIONS PER DAY, CHOLESTEROL, BMI, AGE, SEX, SMOKERLAST 5YRS and EXERCISE MIN PER WEEK. The decision attribute is HEART FAILURE. Out of 10,800 sample values of the dataset, one group consisted of 90% ofthe sample values (9720 sample records) of the dataset using for the learning (training purpose) and another group consisted of 10% of the sample values (1080 sample records) of the dataset for testing purposes.

4.2 RESULTS AND DISCUSSION

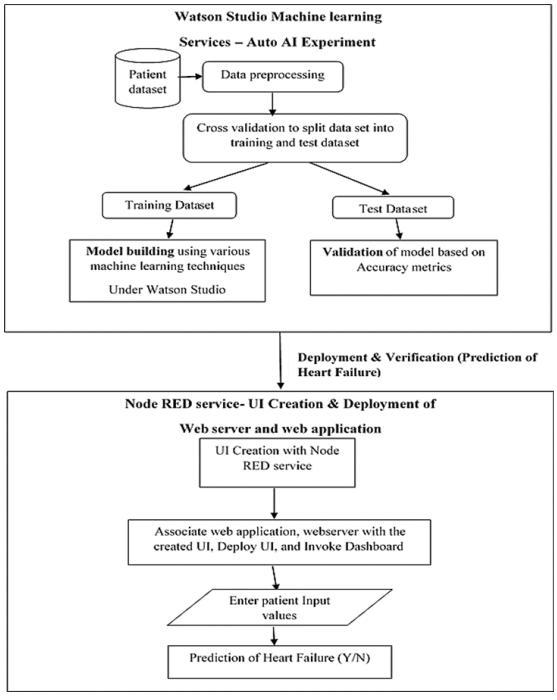
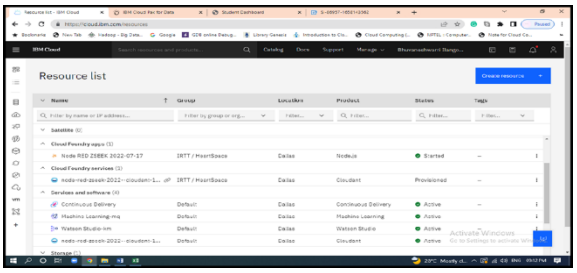


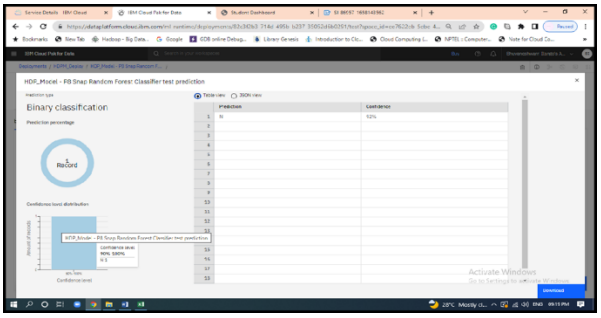
Figure.2 Flow work of Effective Heart Disease Prediction Model using IBM Auto AI

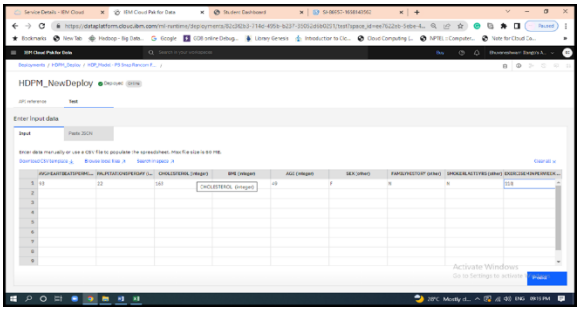
Before the model development, the data preprocessing procedures such as missing value analysis, smoothing

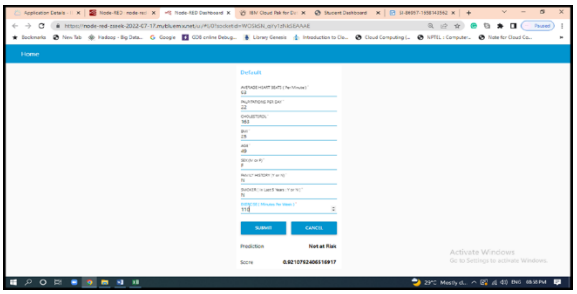
noisy data and data standardization were applied on the patient data set to produce reliable data. Then using cross validation technique, the dataset is split into training dataset and test dataset. The training dataset is used for model development purpose and 10% of the dataset is used for model validation purpose. Then the different prediction models are developed and best prediction model, snap random forest classifier is selected using IBM Auto AI service. The AutoAI graphical tool in Watson Studio automatically analyses patient dataset and generates candidate model pipelines customized for predictive modelling problem. These model pipelines are created iteratively as AutoAI analyses the patient dataset and discovers data transformations, algorithms, and parameter settings that work best for problem setting. Results are displayed on a leader board, showing the automatically generated model pipelines ranked according to the given problem optimization objective. Then this prediction model is validated using test data. The accuracy of the best prediction model obtained is 87.3%. Once the pipeline creation is complete, view and compare the ranked pipelines in a leader board. The Saved model from the action menu for the pipeline with the highest accuracy or low error rate has chosen. This saves the pipeline as a Machine Learning asset in this project. A notification is received thatthe link to view the saved model in this project.Then this model is deployed for ready to use. The API key and the relevant endpoint URLis generated. The user interface is created using IBM node RED service with Node RED flow editor. The Node RED flow in json format is also deployed and associated with prediction model already created. Then invoke the Input User Interface screen with dashboard. Now the patient input details are given through UI and the predicted output is displayed with score. The flow work of this project is clearly depicted in Figure.2.

4.3 SAMPLE SCREEN SHOTS









CONCLUSION

The proposed effective heart disease prediction model is used to predict the HEART FAILURE target attribute of the patient using the following conditional attributes such as AVG HEARTBEATS PER MIN, FAMILY HISTORY, PALPITATIONS PER DAY, CHOLESTEROL, BMI, AGE, SEX, SMOKER LAST 5YRS, EXERCISE MIN PER WEEK of the Patient. The proposedmodel is built with IBM Watson Studio, Node-RED service, Auto AI service, Cloud Object Storage service (COS) and Machine Learning Service. The prediction model is developed with Snap Random Forest Classifier which has accuracy of 87.3%.

FUTURE SCOPE

The model creation, validation and deployment have taken lots of procedures and steps. The aim of the future work is to predict the target attribute by

reducing the number of procedures and steps. The accuracy of the model is also somewhat less compared to already existingprediction models discussed under literature review. In order to improve the accuracy, pipeline structure and algorithm selection procedure will need to be optimized.

REFERENCES

[1] Ramalingam VV, Dandapath A, Raja MK. Heart disease prediction using machine learning techniques: a survey. International Journal of Engineering & Technology. 2018 Oct;7(2.8):684-7.

[2] Mohan S, Thirumalai C, Srivastava G. Effective heart disease prediction using hybrid machine learning techniques. IEEE access. 2019 Jun 19;7:81542-54.

[3] Patel J, TejalUpadhyay D, Patel S. Heart disease prediction using machine learning and data mining technique. Heart Disease. 2015 Sep;7(1):129-37.

[4] Jagtap A, Malewadkar P, Baswat O, Rambade H. Heart disease prediction using machine learning. International Journal of Research in Engineering, Science and Management. 2019 Feb;2(2):352-5.

[5] Khourdifi Y, Bahaj M. Heart disease prediction and classification using machine learning algorithms optimized by particle swarm optimization and ant colony optimization. International Journal of Intelligent Engineering and Systems. 2019 Feb;12(1):242-52.

[6] Gavhane A, Kokkula G, Pandya I, Devadkar K. Prediction of heart disease using machine learning. In2018 second international conference on electronics, communication and aerospace technology (ICECA) 2018 Mar 29 (pp. 1275-1278). IEEE.