Effective Heart Disease Prediction Using IBM Auto AI Service

Submitted

By

Dr. Megha Sharma

1 INTRODUCTION

1.1Overview

Cardiovascular disease (CVD) is a general term for conditions affecting the heart or blood vessels. In this condition, fatty deposits are built inside the arteries (atherosclerosis) and an increased risk of blood clots. As per World Health Organization (WHO), Cardiovascular diseases (CVDs) are the number one cause of death globally, with an estimated 17.9 million people dying from CVDs in 2019, representing 32% of all global deaths. Of these deaths, 85% were due to heart attack and stroke. Heart attacks and strokes are mainly caused by a blockage that prevents blood from flowing to the heart or brain.

There are many disorders associated with the CVDs, which include:

- Coronary heart disease – a disease of the blood vessels supplying the heart muscle;
- Cerebrovascular disease – a disease of the blood vessels supplying the brain;
- Peripheral arterial disease – a disease of blood vessels supplying the arms and legs;
- Rheumatic heart disease – damage to the heart muscle and heart valves from rheumatic fever, caused by streptococcal bacteria;
- Congenital heart disease – congenital impairment that affect the normal development and functioning of the heart caused by malformations of the heart structure from birth; and
- Deep vein thrombosis and pulmonary embolism – blood clots in the leg veins can dislodge and move to the heart and lungs.

Heart attack and stroke are the significant causes of death, which is the focus of this study.

1.2Purpose

This project aims to predict mortality by heart failure using nine features. An unhealthy diet, physical inactivity, tobacco use, and harmful use of alcohol are the most important behavioral risk factors for heart disease and stroke. Behavioral risk factors lead to raised blood pressure, blood glucose, blood lipids, and overweight and obesity. There are also several underlying determinants of CVDs. In this project, we will build a model using Auto AI and a web application where we can showcase the prediction of heart failure.

The following parameters are considered to build a machine learning model that

predicts heart failure:

- Average heart beats per minute (AVGHEARTBEATSPERMIN)
- Palpitations per day (PALPITATIONSPERDAY)
- Cholesterol (CHOLESTEROL)
- Body Mass Index (BMI)
- Age of the individual
- Sex of the individual
- Family history of CVDs (FAMILYHISTORY)
- Is the individual smoker or not (SMOKERLAST5YRS)
- Exercise minutes per week (EXERCISEMINPERWEEK)

2 LITERATURE SURVEY

2.1Existing Problem

Data science is a pool of data operations. In today's world, a substantial amount of information/data is available, which can be used for future predictions. Predictive modeling is a part of data science. Earlier, the problem with prediction was the availability of the tools and technologies to deal with growing volumes and data types. In this study, we will use a machine learning tool, an Auto AI, to predict the chances of heart failure in a person based on their lifestyle.

2.2Proposed Solution

With more interest in using data to produce valuable insights, the availability of faster, cheaper computers, and easier-to-use software prediction, the use of data science is gaining importance. Data science is used for finding patterns, analyzing, and deriving inferences within the data, which are derived using several statistical methodologies. In data science, examining the data from data extraction to wrangling and pre-processing is done, which helps develop data-driven forecasts.

The goal here is to extract conclusions from data, aiding firms in making better business decisions. Predictive modeling is an essential part of Data Science. It generates predictions based on historical data. We can get in-depth knowledge of businesses, organizations, and more using predictive modeling. Organizations are turning to predictive analytics to help solve challenging problems and uncover new opportunities. Predictive modeling makes use of statistics to forecast the outcomes and machine learning.

In this study, we will predict heart failure using the historical data set using IBM services like Watson Studio, Watson Machine Learning, Node-RED, and Cloud Object Storage.

3 THEORETICAL ANALYSIS

3.1Block diagram

Figure 1 shows the flowchart illustrating the procedure adopted for this study. The data set collected in 'csv' format contains ten columns and 10800 rows, with information collected on nine parameters and heart failure records. This data set is used for developing a machine learning model for heart failure prediction.

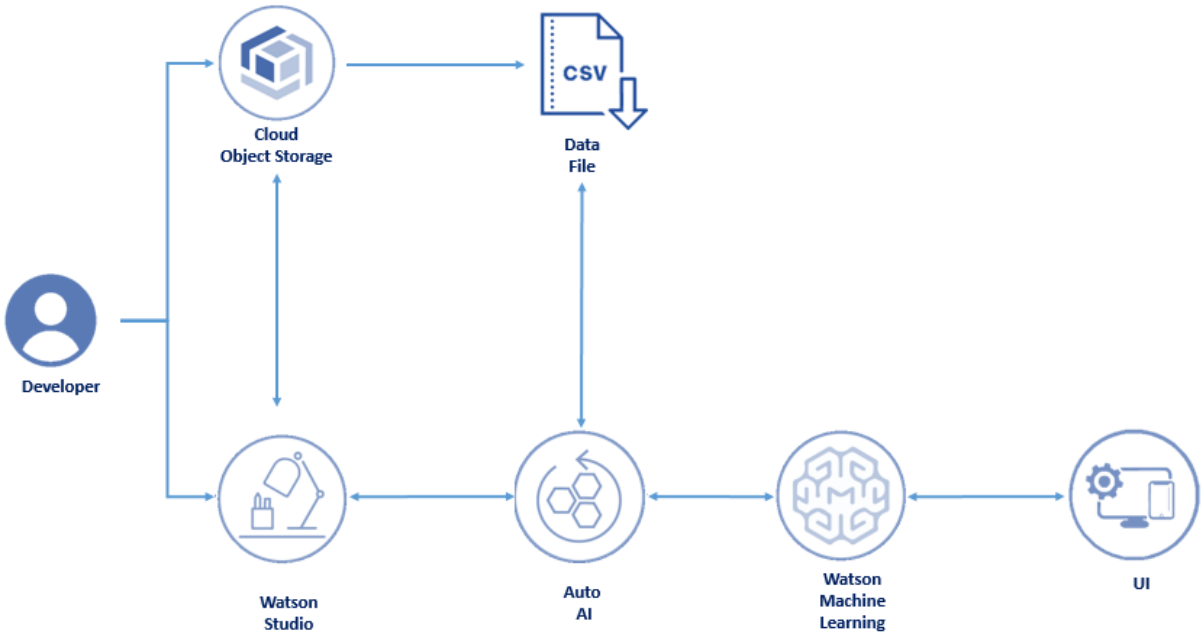


Figure 1 Overview of the Project.

The methodology adopted for this project is as follows:

1. Create IBM Watson Studio and Node-RED Service
2. Create a Watson studio project and add the Auto AI Experiment
3. Run the Auto AI Experiment to build a Machine learning model on the desired dataset
4. Save and deploy the model as a web server and generate a scoring endpoint
5. Create a WEB application using Node-RED to take user input and showcase predictions on user interface (UI)

3.2Software Designing

The IBM services used in this project are:

1. IBM Watson Studio is IBM's software platform for data science. It consists of multiple collaboration and open-source tools useful in data science. We can create a project using Watson Studio and work with various analytical tools and languages.
2. IBM Watson machine learning service is a full-service IBM Cloud offering that makes it easy for developers and data scientists to work together to integrate predictive capabilities with their applications. The Machine Learning service is a set of REST APIs you can call from any programming language to develop applications that make smarter decisions, solve challenging problems, and improve user outcomes.
3. Node-Red is a programming tool for wiring together hardware devices, APIs, and online services. Primarily, it is a visual tool designed for the Internet of Things, but it can also be used for other applications to assemble various services' flows quickly. Node-RED enables users to stitch together Web services and hardware by replacing common low-level coding tasks (like a simple service talking to a serial port). This can be done with a visual drag-drop interface. Various components in Node-RED are connected to create a flow. Most of the code needed is created automatically.
4. Auto AI is a graphical tool in Watson Studio that automatically analyzes the data and generates candidate model pipelines customized for the predictive modeling problem. These model pipelines are created iteratively as AutoAI investigates your dataset and discovers data transformations, algorithms, and parameter settings that work best for the problem set. Results are displayed on a leaderboard, showing the automatically generated model pipelines ranked according to the problem optimization objective.

4 EXPERIMENTAL INVESTIGATION

While working on this project, the first step was to create an IBM cloud account and add the IBM Watson studio service. Further, create a project to predict heart failure using the dataset provided. Upload the data set to the Watson studio project. Next, create an Auto AI experiment to create a model. We must build a machine learning instance while creating the Auto AI experiment. Next, deploy the machine learning model as a web server and integrate it with Node-red Service. Next, create a user interface (UI) with a node-red application. Once the machine learning model is deployed, the prediction analysis is done using a different algorithm, as shown in Figure 2.

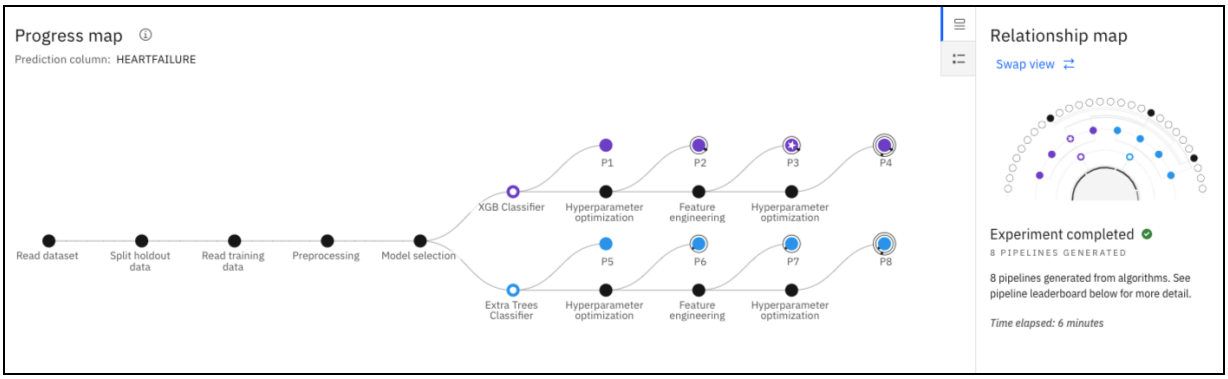


Figure 2 Progress of Heart Failure Prediction.

Figure 3 illustrates the accuracy and prediction of eight pipelines used in machine learning to come up with the pipeline which has higher accuracy.

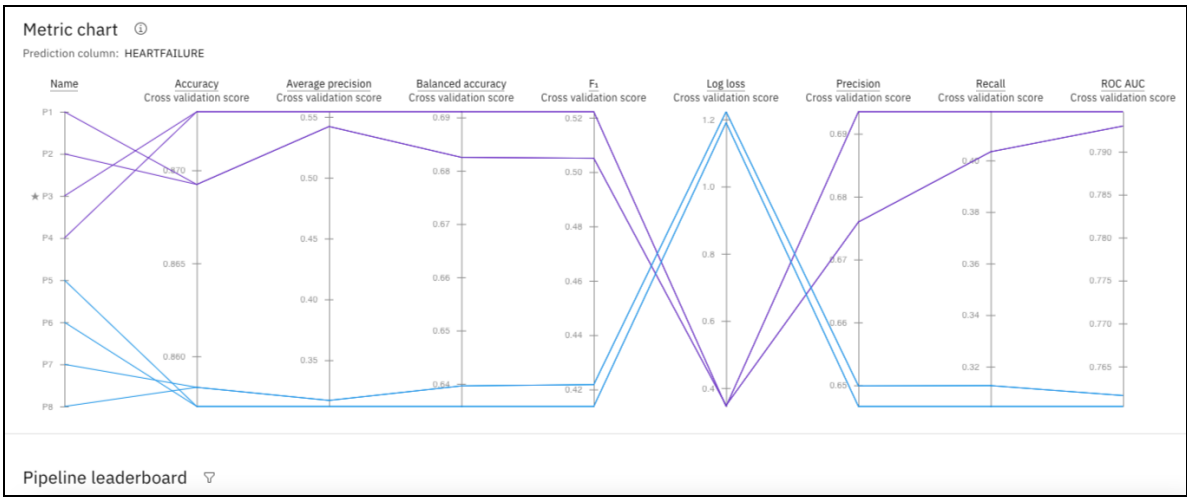
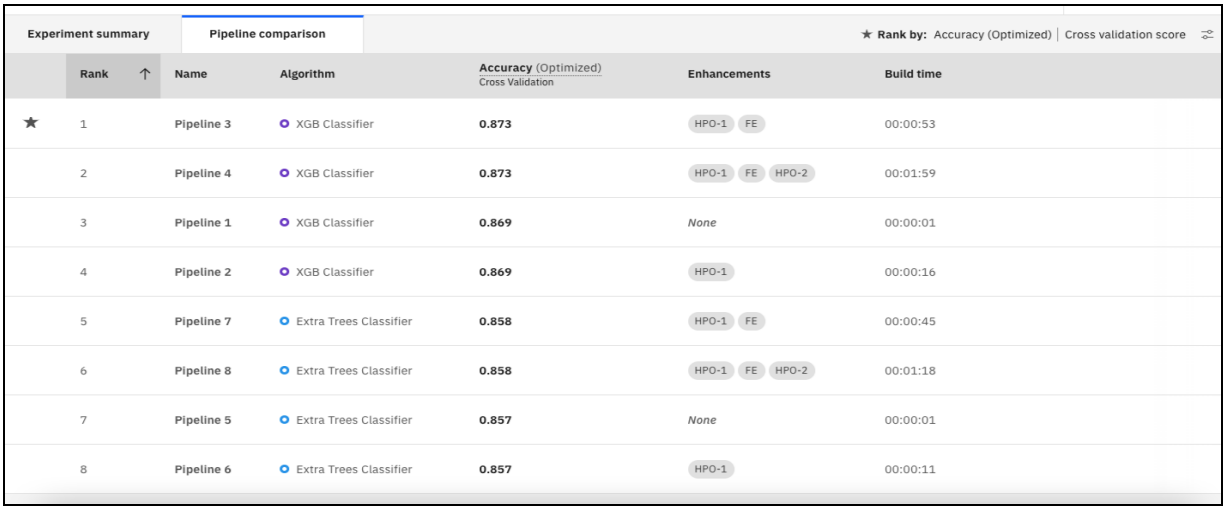


Figure 3 Pipeline Comparison.

Eight pipelines were based on the XGB classifier and Extra trees classifier, their accuracy, enhancements, and build time are shown in Figure 4. Based on this information, Pipeline 3, with the highest accuracy of 0.873, was adopted for this study.



Experiment summary | Pipeline comparison | ★ Rank by: Accuracy (Optimized) | Cross validation score

	Rank	Name	Algorithm	Accuracy (Optimized) Cross Validation	Enhancements	Build time
★	1	Pipeline 3	XGB Classifier	0.873	HPO-1 FE	00:00:53
	2	Pipeline 4	XGB Classifier	0.873	HPO-1 FE HPO-2	00:01:59
	3	Pipeline 1	XGB Classifier	0.869	None	00:00:01
	4	Pipeline 2	XGB Classifier	0.869	HPO-1	00:00:16
	5	Pipeline 7	Extra Trees Classifier	0.858	HPO-1 FE	00:00:45
	6	Pipeline 8	Extra Trees Classifier	0.858	HPO-1 FE HPO-2	00:01:18
	7	Pipeline 5	Extra Trees Classifier	0.857	None	00:00:01
	8	Pipeline 6	Extra Trees Classifier	0.857	HPO-1	00:00:11

Figure 4 Details of the Pipeline adopted for the Prediction.

Figure 5 shows the percentage of training data and holding data for testing.

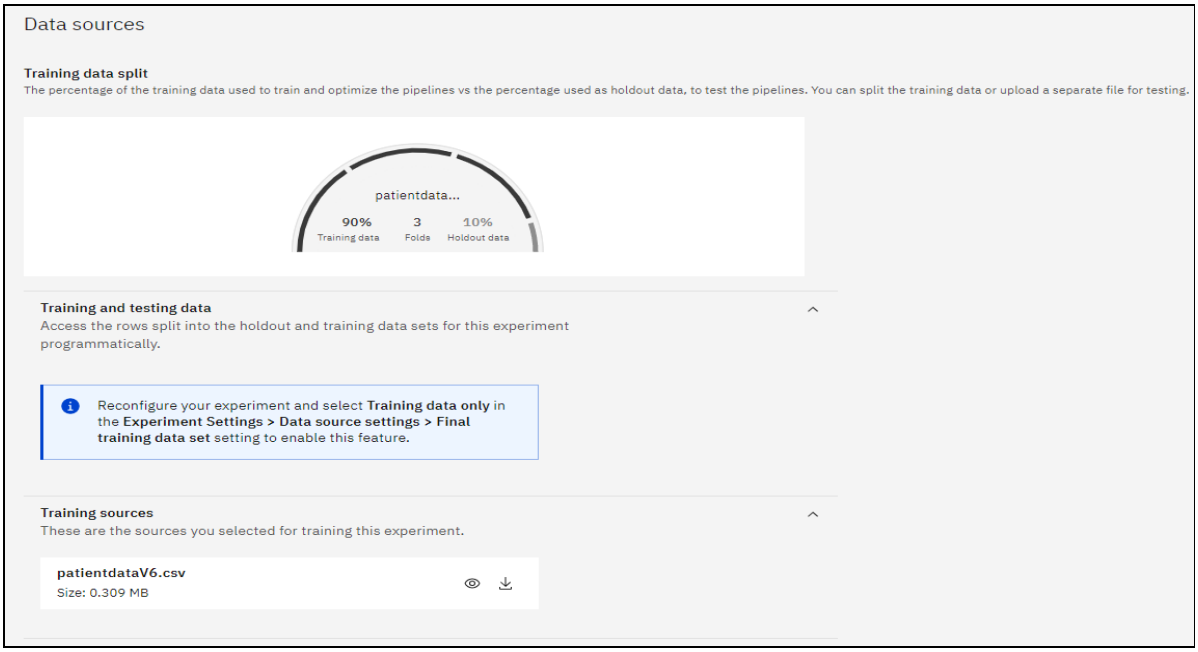


Figure 5 Details of Data source predictions.

5 RESULTS AND DISCUSSION

Figure 6 illustrates the interface of the Node-Red Application, where flows can be dragged and added to the desired use.

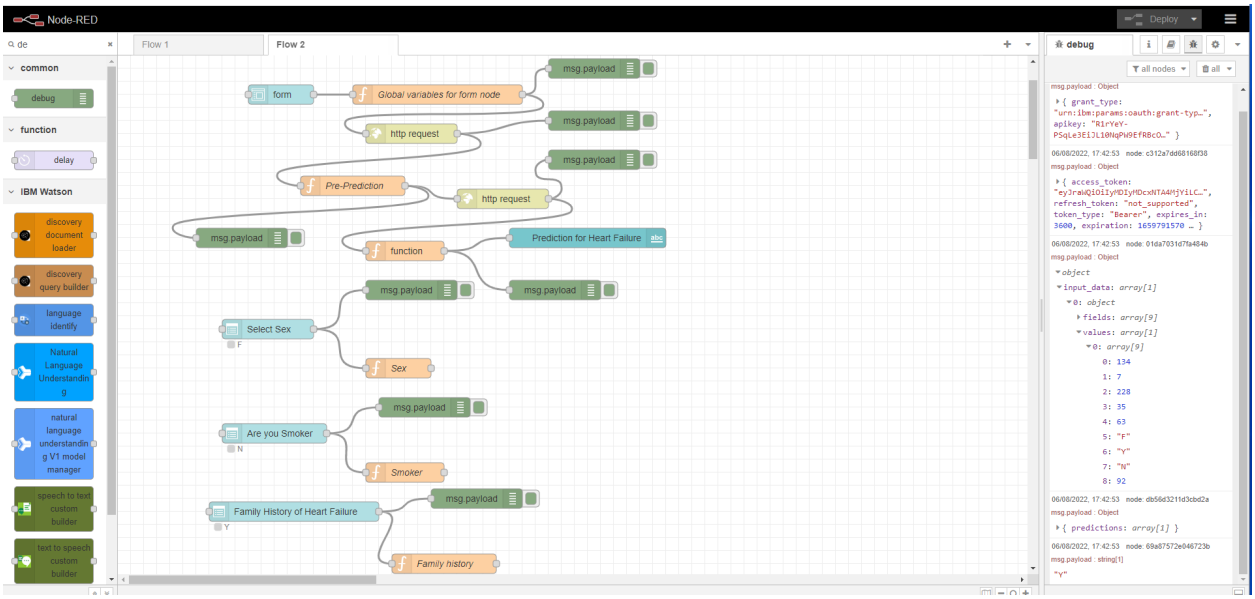


Figure 6 Node-Red Interface.

In node-red, using the steps mentioned, a user interface (UI) was created:

1. Create the UI by dragging form nodes and debug nodes, as noted in Figure 7
2. Grab the values using user input after making the dashboard prediction, as shown in Figure 8.
3. Next, we need to integrate the machine learning scoring endpoint into the UI by first setting the variables global.
4. Next, get the access token with the help of an HTTP request.
5. Afterward, send the input variables to the scoring endpoint and the access token.

Get the predicted value and showcase it on the UI (Figure 8). An example prediction is shown in the dashboard figure.

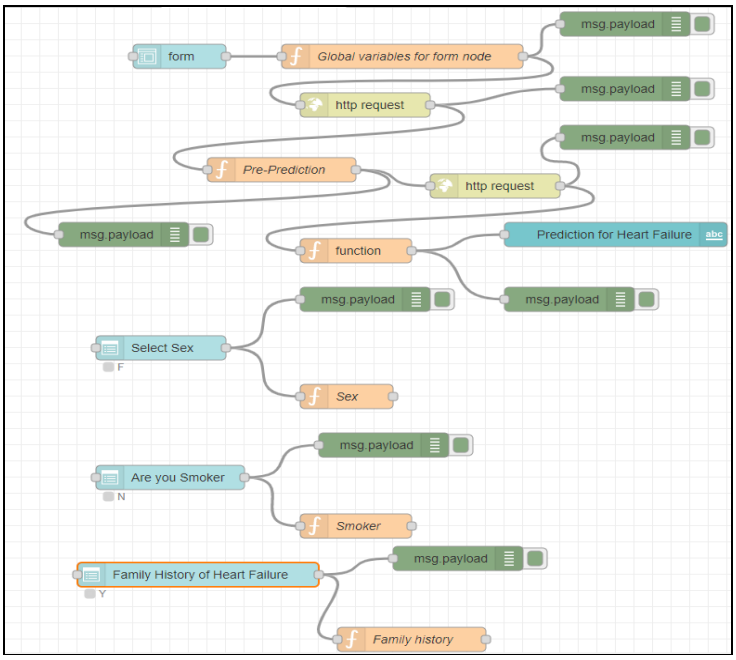


Figure 7 Node-Red Flows.

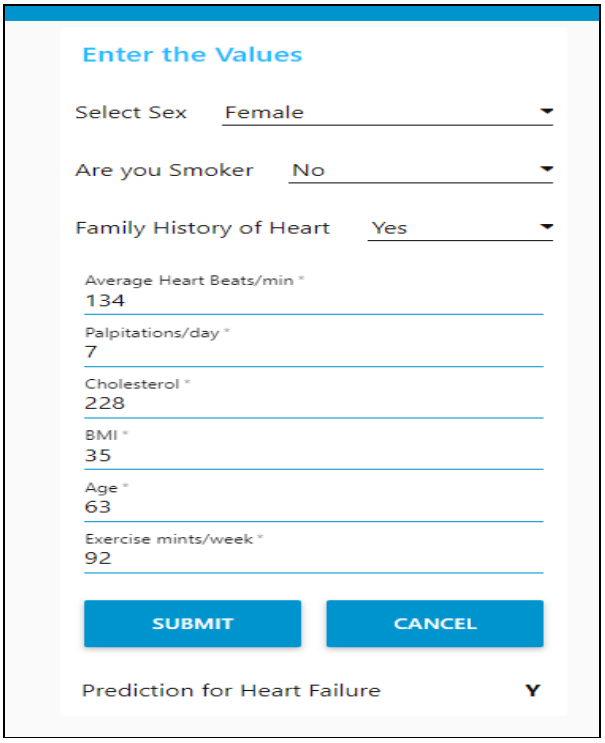


Figure 8 Node-Red UI.

Node Red Dashboard is accessible at: https://node-red-hnvbn-2022-07-22.mybluemix.net/ui/#!/0?socketid=gqZkKtiyNVJ7BV39AAAh

The API reference is: https://us-south.ml.cloud.ibm.com/ml/v4/deployments/d566edea-ac71-415b-93b2-ed9ee99a1f39/predictions?version=2022-07-22

To import the deployed machine learning prediction model following python code can be used:

```
# NOTE: you must manually set API_KEY below using information retrieved from your IBM Cloud account.

API_KEY = "<your API key>"

token_response = requests.post('https://iam.cloud.ibm.com/identity/token', data={"apikey":

 API_KEY, "grant_type": 'urn:IBM:params:oauth:grant-type:apikey'})

mltoken = token_response.json()["access_token"]


header = {'Content-Type': 'application/json', 'Authorization': 'Bearer ' + mltoken}


# NOTE: manually define and pass the array(s) of values to be scored in the next line

payload_scoring = {"input_data": [{"fields": [array_of_input_fields], "values": [array_of_values_to_be_scored, another_array_of_values_to_be_scored]}]}


response_scoring = requests.post('https://us-south.ml.cloud.ibm.com/ml/v4/deployments/d566edea-ac71-415b-93b2-ed9ee99a1f39/predictions?version=2022-07-22', json=payload_scoring,

 headers={'Authorization': 'Bearer ' + mltoken})

print("Scoring response")

print(response_scoring.json())
```

6 ADVANTAGES & DISADVANTAGES

Using machine learning the whole prediction analysis became easy and fast. The Using machine learning, the whole prediction analysis became easy and fast. The advantages of using ML in this project for predicting Heart Failure are as follows:

1. **Easily identifies trends and patterns:** Using Machine Learning large volume of heart failure data was reviewed, and specific trends and patterns were identified.
2. **No human intervention needed (automation)**: With ML, in a matter of minutes, eight pipelines based on different algorithms were analyzed.
3. **Continuous Improvement:** ML algorithms keep improving and select the one with higher accuracy and efficiency.

Some factors limiting the ML are data acquisition, time, and resources in case of a large amount of data to be reviewed. Choosing the best algorithm for interpreting the results sometimes limits the ML. Small datasets that are not inclusive can lead to biases and prediction errors.

7 APPLICATIONS

With predictive analytics using ML, we can discover insights about the future. Areas in which this is applicable are:

- Banking and Financial Services
- Retail
- Oil, Gas & Utilities
- Governments & the Public Sector
- Health Insurance
- Manufacturing

8 CONCLUSIONS

In conclusion, the first thing to start with prediction in ML we need a problem to solve and understand what we need to know and predict. Understand the need and decisions to be taken based on the data provided. As for the present case, we can work on improving the conditions after considering the parameters and finding the factors which lead to heart failure. Using the UI interface to prediction is easier to understand and implement.

9 BIBILOGRAPHY

1. https://www.sas.com/en_us/insights/analytics/predictive-analytics.html
2. https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)
3. https://cloud.ibm.com/catalog/services/machine-learning#about
4. Advantages and Disadvantages of Machine Learning Language - DataFlair (data-flair.training)