Effective Heart Disease Prediction Using IBM Auto AI Service

1.1 Overview

Cardiovascular diseases (CVDs) are the number 1 cause of death globally, taking an estimated 17.9 million lives each year, which accounts for 31% of all deaths worldwide.
Heart failure is a common event caused by CVDs and this dataset contains 9 features that can be used to predict mortality by heart failure.
In this project, i have build a model using Auto AI and build a web application where i can showcase the prediction of heart failure.

1.2 Purpose

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.

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 FunctionTransductive 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 optimizationTalos 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 is found to be accurate by 90.76%.

2.2 Proposed solution

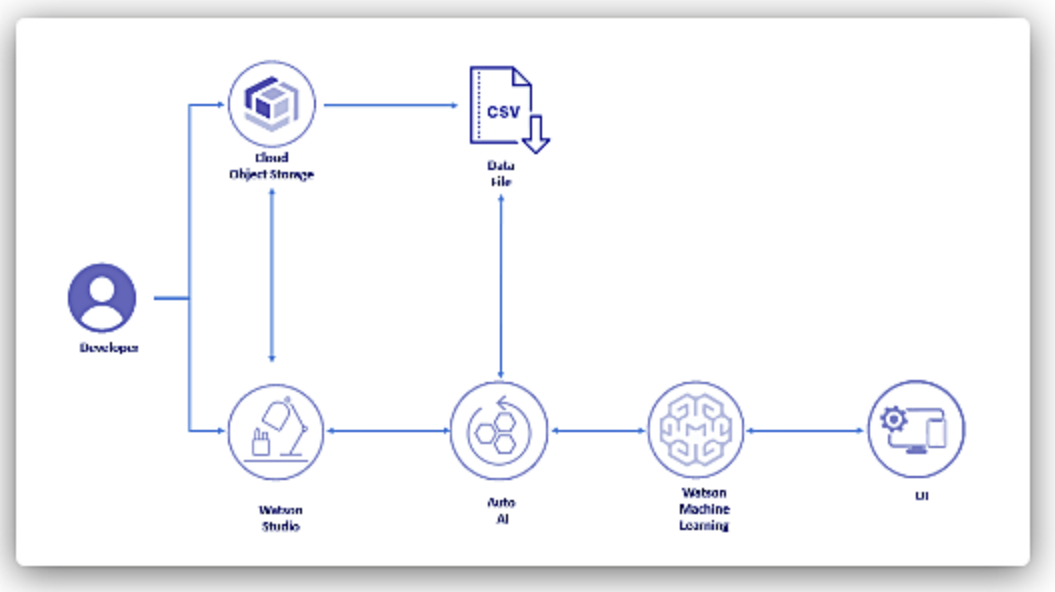
The main focus of this system is to automate the prediction of heart failure by using the AutoAI service where separate data preprocessing, feature extraction steps and development need not be processed. Auto AI automates Data preparation, Model development, Feature engineering and Hyper-parameter optimization. The service is available in IBM Watson Studio[8]. This makes data scientists to quickly get started and expert data scientists to accelerate 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 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[9] used is a repository from IBM used for the purpose of heart failure prediction. It has 9 columns with 10801 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 produces an 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.

Theoretical ANALYSIS
People with heart disease tend to be older and male, and have higher blood pressure, higher cholesterol levels, lower maximum heart rate under the Thalium stress test, etc., than people without the disease.

3.1 Block diagram

The developer creates a Cloud Pak for Data project.

1. A model is created with AutoAI by uploading some data.
2. Data is backed up and stored on Cloud Object Storage.
3. The model is deployed using the Watson Machine Learning service.
4. A Node.js web app is deployed on IBM Cloud. It calls the predictive model hosted on the Watson Machine Learning service.
5. A user visits the web app, enters their information, and the predictive model returns a response.



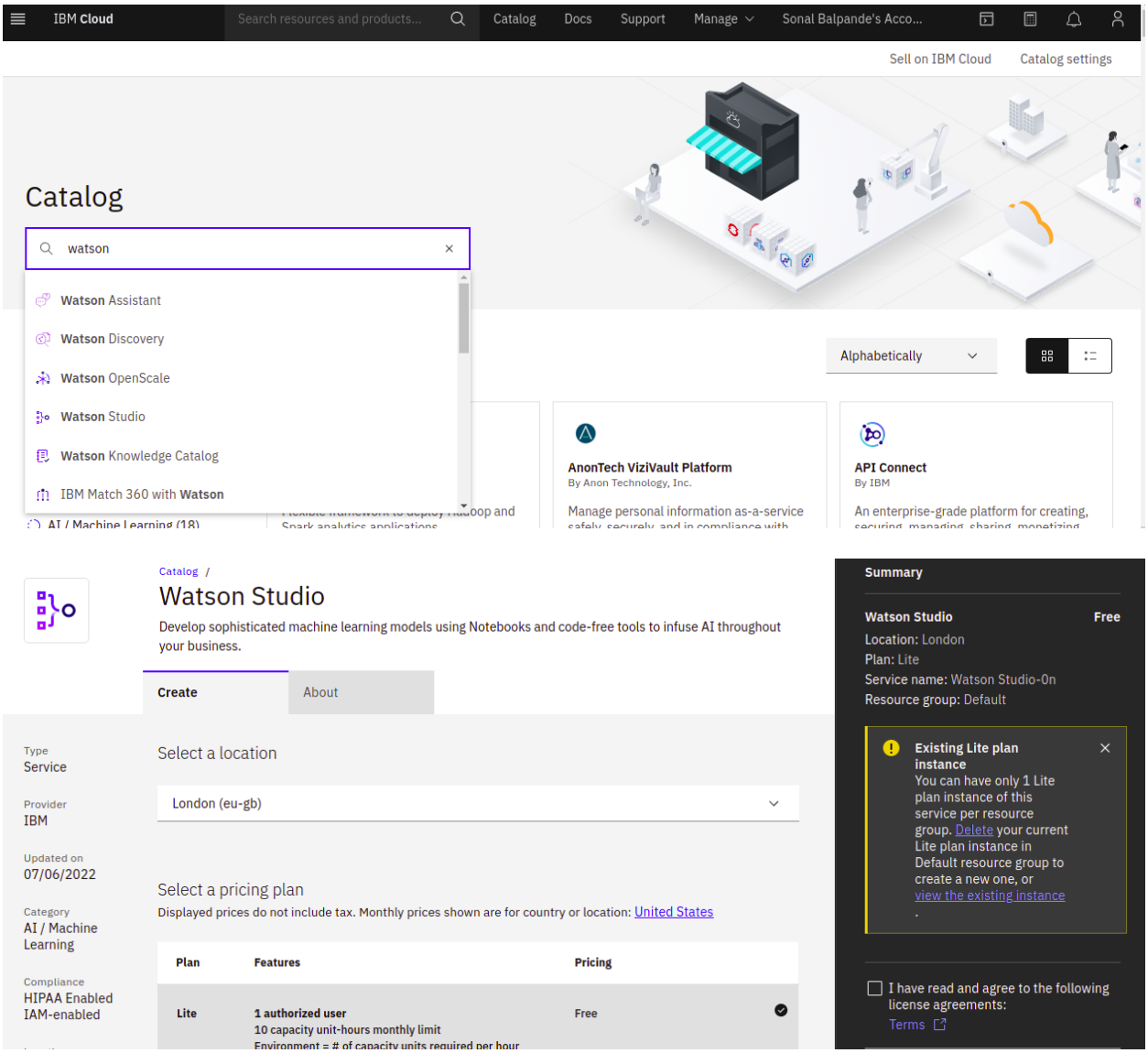
Diagrammatic overview of the project.

3.2 Hardware / Software designing

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

4 EXPERIMENTAL INVESTIGATIONS

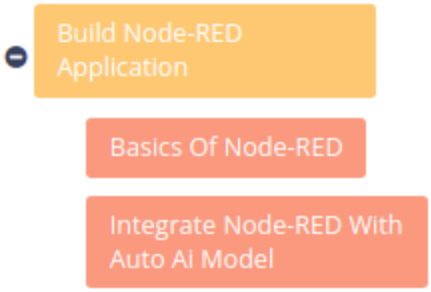
Experiment consist of following steps
Create IBM services consist of
a)Create Watson studio
Click on Catalogue and search Watson Studio.Select Watson Studio



Create instance of Watson Studio

b)Create Node Red Service

Goto Catalogue ,select Software and search for Node-Red and select Node-Red App. Create and than deploy the app.Select Cloud Foundry .Create API key.

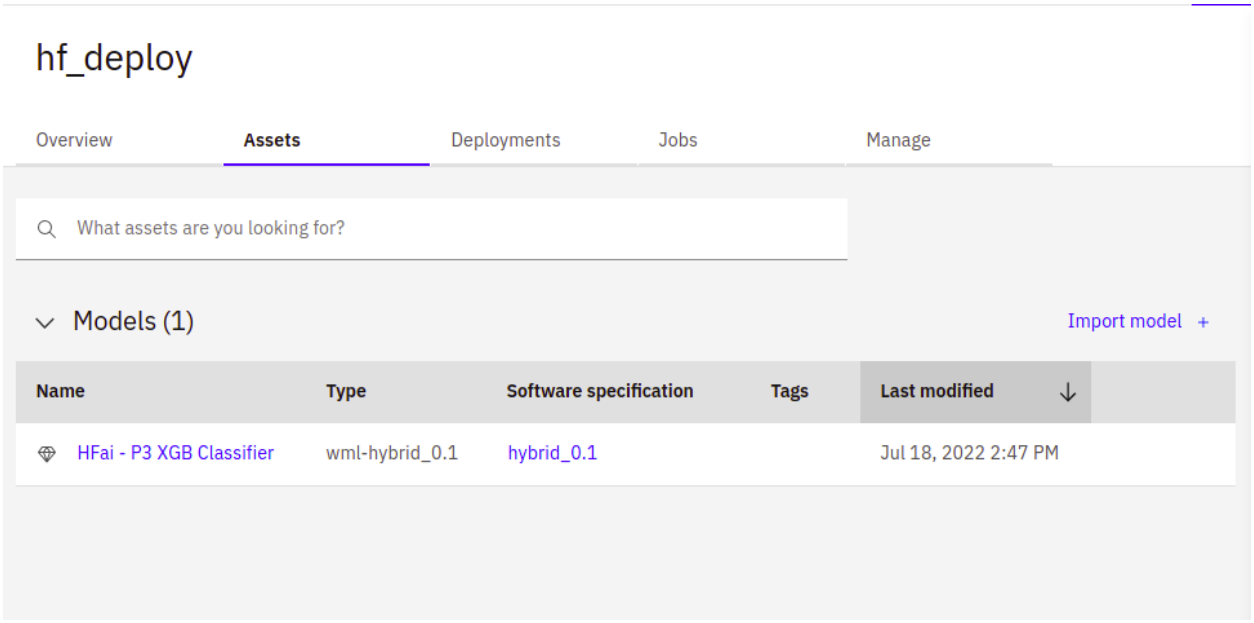


Build Machine Learning Model

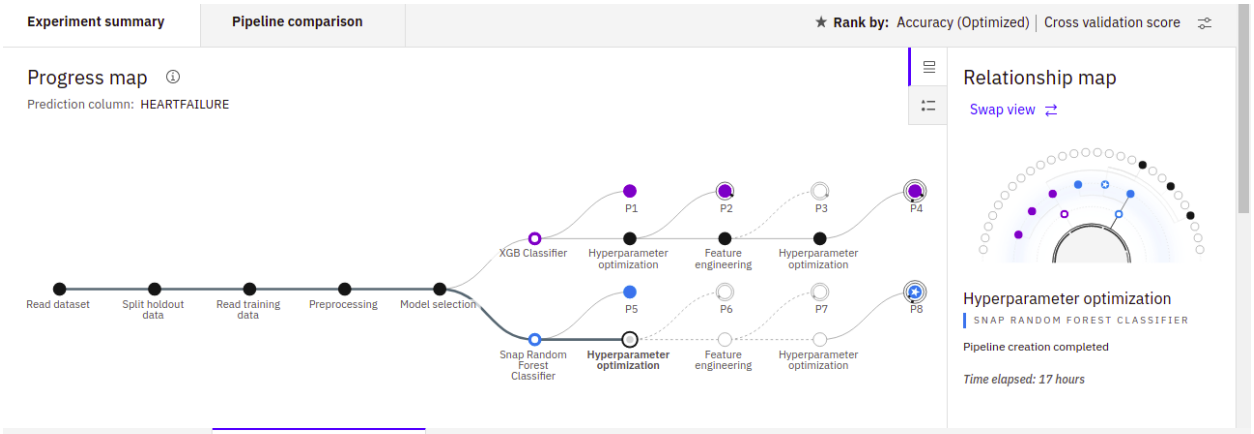
In this activity, you will learn how you can save a pipeline as a Watson Machine Learning model, deploy the model, and score it to view a prediction.
This Activity contains the Following Tasks

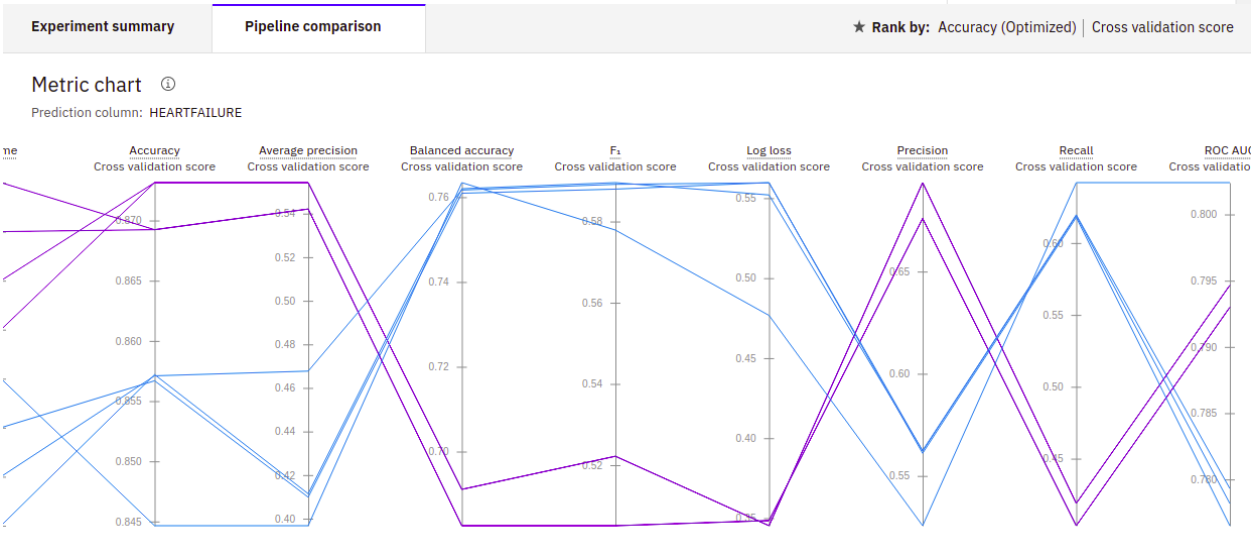
- Collect the data set
- Create Watson Studio project
  Create new Project, Add Project Name
- Create a new project by clicking Create a project.
- Choose an Empty project.
- Enter a Name and associate the project with a Cloud Object Storage service.
- At the project dashboard click on the Assets tab and upload the data set associated with this repo. patientdataV6.csv
- Add Auto AI experiment
  - ➤ Start by clicking on Add to project and choosing AutoAI experiment.
  - ➤ Give it a Name and specify a Watson Machine Learning instance.
  - ➤ Choose to use data from your project.
  - ➤ Choose the patientdataV6.csv option.
  - ➤ For the "What do you want to predict?" option, choose HEARTFAILURE.
  - ➤ he experiment will take a few minutes to run.
  - ➤ Once completed hover over the top option to make the Save as button appear. Click it.
  - ➤ Choose to save the experiment as a Model.
  - ➤ Download a generated Jupyter Notebook that can be used to re-create the steps..
  - ➤ That were taken to create the model
  - ➤ You model will be saved. Click the dialog to view it in your project.
  - ➤ Once you're at the model overview choose the button Promote to deployment space.

- Run AI Experiment
- Save the model
- Deploy the model
  - ➤ Choose the deploy the model by clicking the rocket ship icon.(On right top corner)
  - ➤ Choose the Online deployment option and give it a name.
  - ➤ Click on the API reference tab and save the Endpoint. We will be using this in our application.

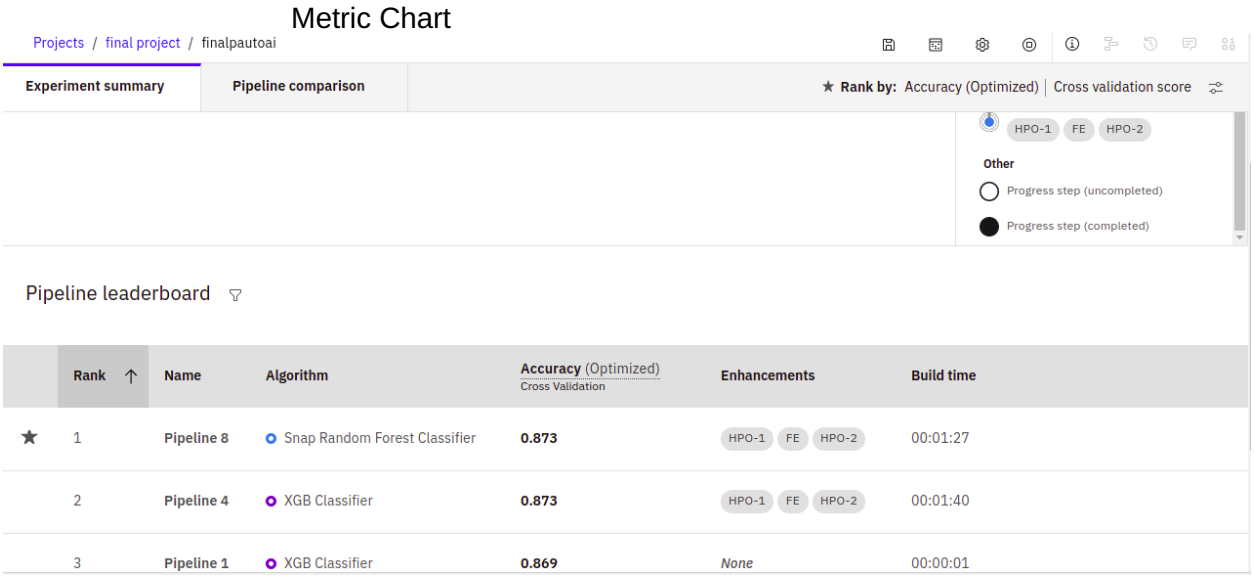

hf_deploy
Overview
Assets
Deployments
Jobs
Manage
What assets are you looking for?
Models (1)
Import model +
Name
Type
Software specification
Tags
Last modified
HFai - P3 XGB Classifier
wml-hybrid_0.1
hybrid_0.1
Jul 18, 2022 2:47 PM

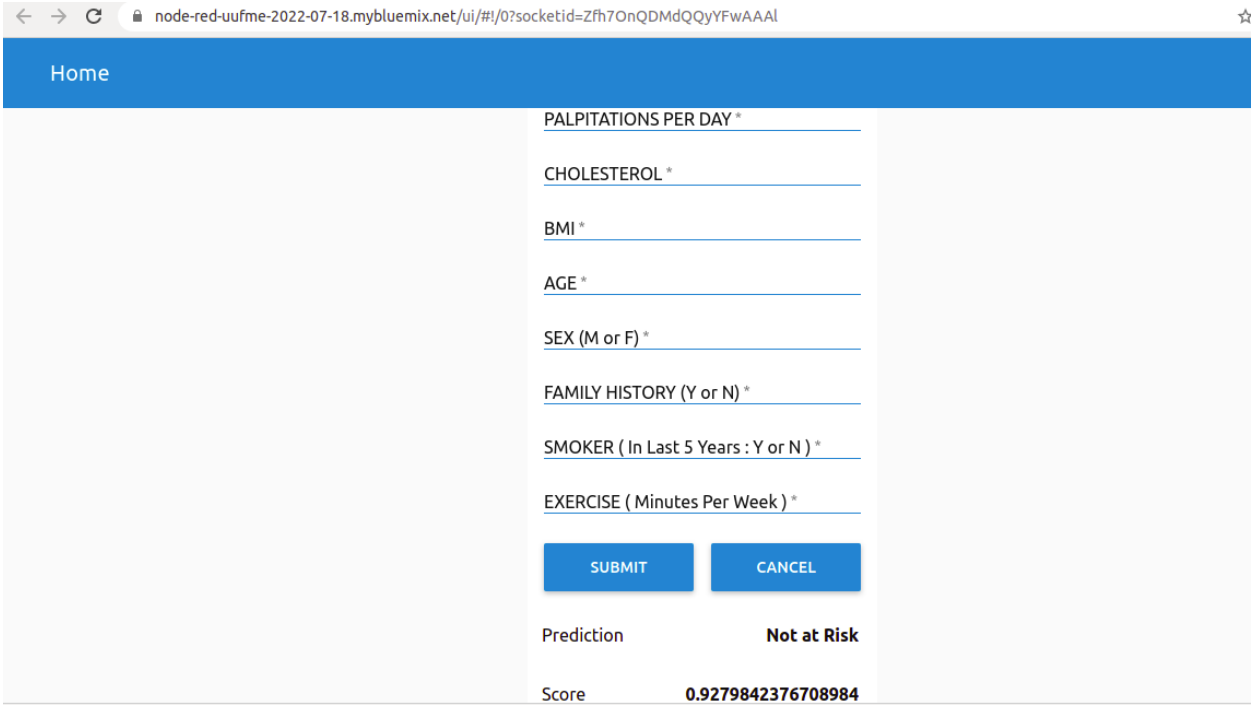
6 RESULT


Experiment summary
Pipeline comparison
★ Rank by: Accuracy (Optimized) | Cross validation score
Progress map
Prediction column: HEARTFAILURE
Read dataset
Split holdout data
Read training data
Preprocessing
Model selection
XGB Classifier
Snap Random Forest Classifier
P1
P2
P3
P4
P5
P6
P7
P8
Hyperparameter optimization
Feature engineering
Hyperparameter optimization
Hyperparameter optimization
Feature engineering
Hyperparameter optimization
Relationship map
Swap view
Hyperparameter optimization
SNAP RANDOM FOREST CLASSIFIER
Pipeline creation completed
Time elapsed: 17 hours


Experiment summary
Pipeline comparison
★ Rank by: Accuracy (Optimized) | Cross validation score
Metric chart
Prediction column: HEARTFAILURE
Accuracy
Cross validation score
Average precision
Cross validation score
Balanced accuracy
Cross validation score
F1
Cross validation score
Log loss
Cross validation score
Precision
Cross validation score
Recall
Cross validation score

Metric Chart





Final findings (Output) of the project along with screenshots.

9 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.
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.

10 FUTURE SCOPE

To obtain a more generalized classification and prediction accuracy, other multiple heart disease datasets from geographically diverse sources with more features should be explored for developing more efficient machine learning models, and that is the fundamental intent of our future research, which is on progress. With this, more efficient classification and early prediction of heart diseases would be achieved, which in turn minimizes the escalating rates of morbidity and mortality due to CVDs