Predict Heart Failure Using IBM Auto AI Service

Introduction

Heart failure is a condition that develops when your heart doesn't pump enough blood for your body's needs. This can happen if your heart can't fill up with enough blood. It can also happen when your heart is too weak to pump properly. The term "heart failure" does not mean that your heart has stopped. However, heart failure is a serious condition that needs medical care.
Heart failure can develop suddenly (the acute kind) or over time as your heart gets weaker (the chronic kind). It can affect one or both sides of your heart. Left-sided and right-sided heart failure may have different causes. Most often, heart failure is caused by another medical condition that damages your heart. This includes coronary heart disease, heart inflammation, high blood pressure, cardiomyopathy,Cardiovascular diseases or an irregular heartbeat . Heart failure may not cause symptoms right away. But eventually, you may feel tired and short of breath and notice fluid buildup in your lower body, around your stomach, or neck.

Purpose

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, go to build a model using Auto AI and build a web application where we can showcase the prediction of heart failure.

Project Objectives

Services Used:

- IBM Watson Studio
- IBM Watson Machine Learning
- Flask Application
- IBM Cloud Object Storage

Build an Application using flask which takes inputs from the user and showcases the prediction on UI. use the correct machine learning model to predict the results.

Literature Survey

A.Existing problem

Work with Watson Studio and Create a project in Watson Studio and then create an auto AI experiment to create a ML model. This auto AI create has some limitations while creating it in auto AI.so pipelines are create from auto AI and deployed the model.probelms like predicting incorrect results,less accuracy.

B.Proposed solution

We gonna build a machine learning model that predicts heart failure based on the following parameters

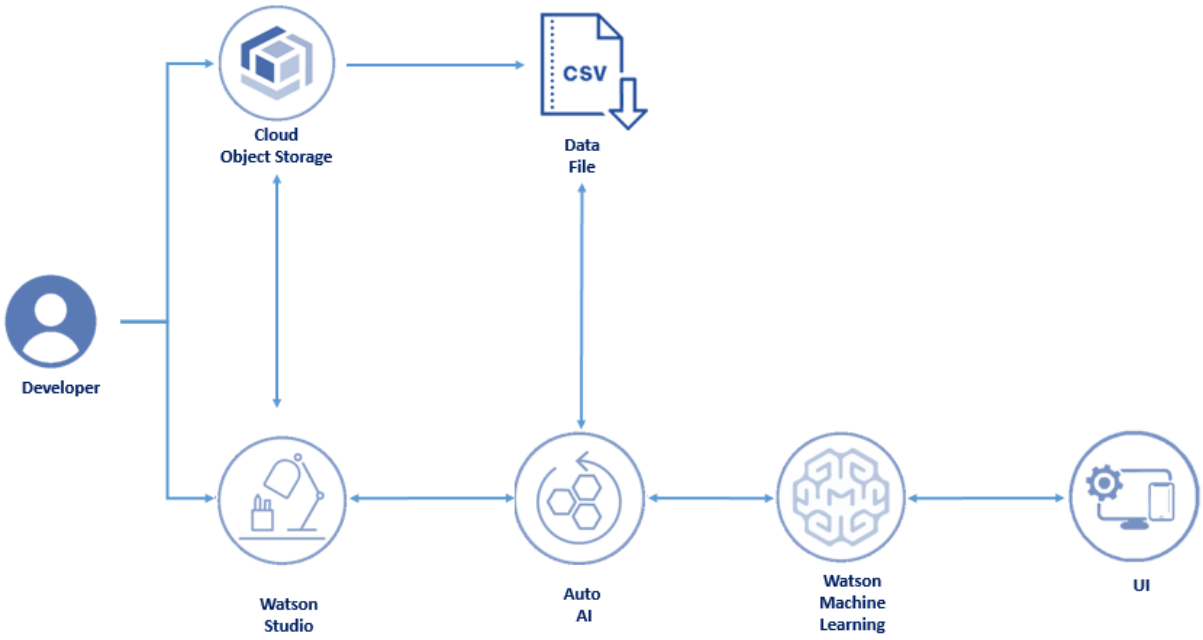
- AVGHEARTBEATSPERMIN
- PALPITATIONSPERDAY
- CHOLESTEROL
- BMI
- AGE
- SEX
- FAMILY HISTORY
- SMOKERLAST5YRS
- EXERCISEMINPERWEEK

we gonna collect the data,clean the data,visualise the data and finally Create a WEB application Using red-node to take user input and showcase Prediction on UI.

Theoretical Analysis

A.Block diagram

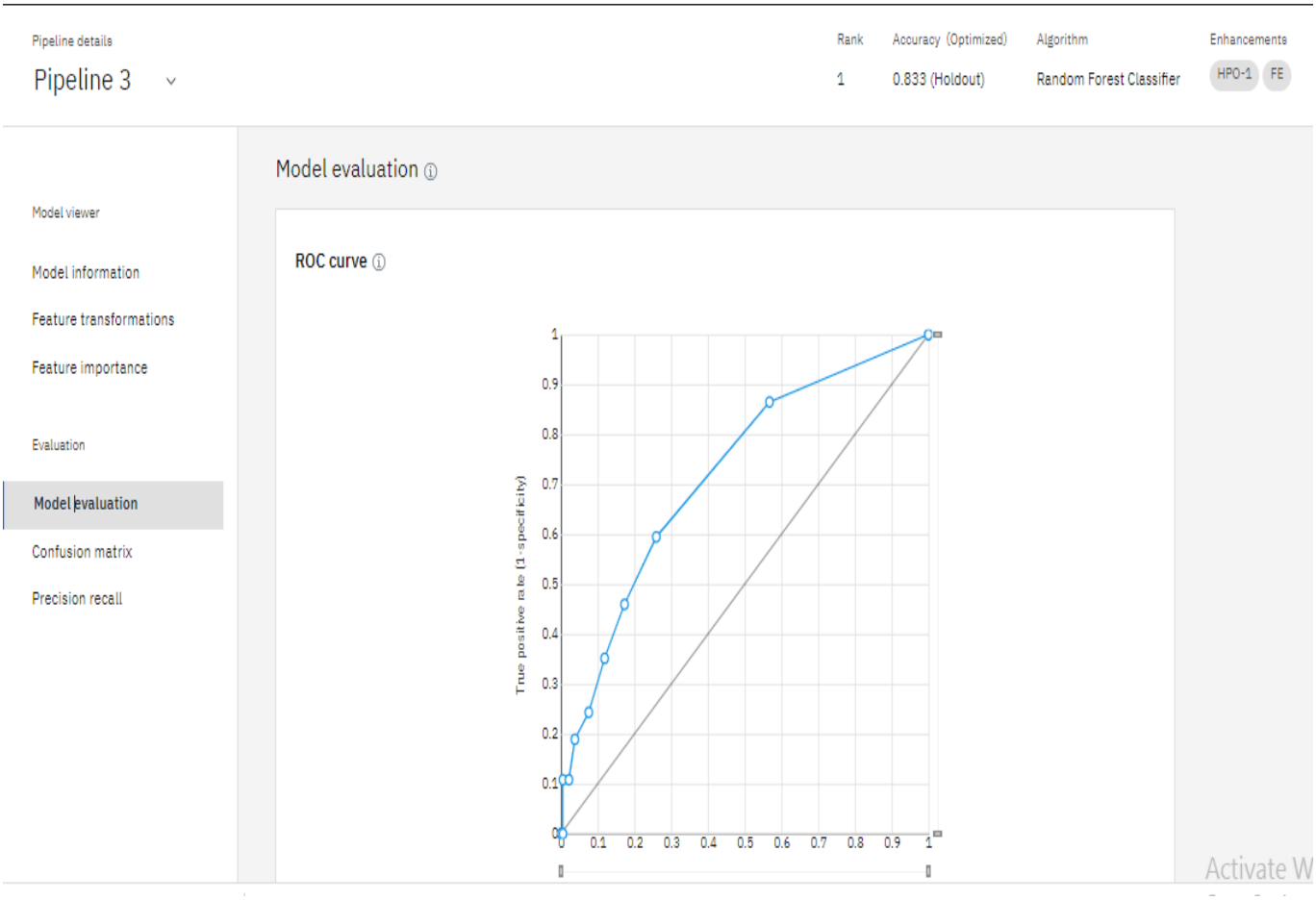
To achieve the project goal we have the steps to follow. As mentioned in block diagram, we developer create a watson studio project, cloud object storage and import data set to the auto AI to get the best pipeline and then create a watson machine and integrated with red-node application to take user input and showcase Prediction on UI.



Experimental Investigations

- Collected the data file from github and import to jupyter notebook for data preprocessing and to train the data.
- Data cleaning is done by checking all the data wrangling steps.
- Since the target data is in binary here we use suitable classifier algorithm to get correct output from the 9 inputs.
- Here the data is unbalanced hence we use oversampling to make the data balanced.
- data visualization is done to know more about the data.
- from data visualization came to know that SEX input has equal is unwanted or same mean value.
- groupby allow to split the categarical data into seperate groups to perform computations for better analysis.
- delete or drop the unwanted column when it is having a same mean value.
- LabelEncoding is used to convert the categorical class into binary.
- LabelEncoding is done in FAMILYHISTORY,SMOKERLAST5YRS ,EXERCISEMINPERWEEK and HEARTFAILURE.

- seperate the input and target data as x and y respectively.
- feature scaling is done to get all the column values ranging from -1 to 1
- train_test_split is done to predict the ML model correct.
- model selection is RandomForestClassifier which gives good results compared to other models.
- hyperparameter tunning is done to predict the best permutation and combination.
- observing the confusion matrix, this RandomForestClassifier model gives the lowest number false negative and false postive.
- finally resolving all the probelms in data set, accuracy_csore is 96%
- The result from the auto AI is 83% and it is increased to 96%.
- confusion matrix with less number of false negative and false positive value.
- In fig shown, accuracy obtained from the auto AI as 83%.



pipeline 3 -RandomForestClassifier-0.833 accuracy

```
In [77]: from sklearn.ensemble import RandomForestClassifier
         rfc=RandomForestClassifier()
         rfc.fit(x_train,y_train)

         C:\Users\ACER\anaconda3\lib\site-packages\ipykernel_launcher.py:3: DataConversionWarning: A column-vector y was passed when a
         d array was expected. Please change the shape of y to (n_samples,), for example using ravel().
           This is separate from the ipykernel package so we can avoid doing imports until

Out[77]: RandomForestClassifier(bootstrap=True, ccp_alpha=0.0, class_weight=None,
                                criterion='gini', max_depth=None, max_features='auto',
                                max_leaf_nodes=None, max_samples=None,
                                min_impurity_decrease=0.0, min_impurity_split=None,
                                min_samples_leaf=1, min_samples_split=2,
                                min_weight_fraction_leaf=0.0, n_estimators=100,
                                n_jobs=None, oob_score=False, random_state=None,
                                verbose=0, warm_start=False)

In [78]: y_test

Out[78]: array([[0],
                [1],
                [0],
                ...,
                [1],
                [0],
                [0]])

In [79]: y_pred=rfc.predict(x_test)
         y_pred

Out[79]: array([0, 1, 1, ..., 1, 0, 1])

In [80]: from sklearn.metrics import accuracy_score,confusion_matrix,classification_report
         accuracy_score(y_pred,y_test)

Out[80]: 0.9614424410540915
```

jupyter notebook output as 0.96 accuracy

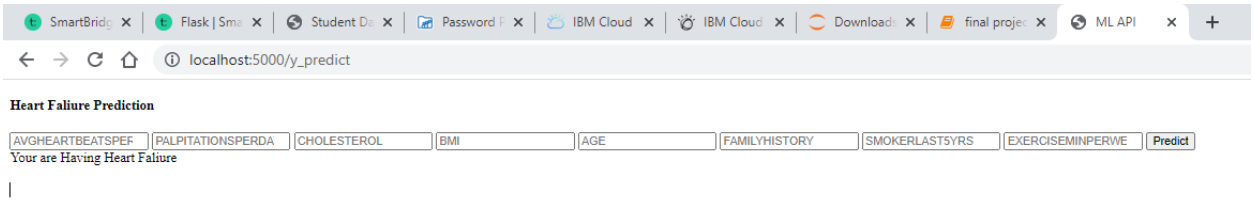
Flowchart

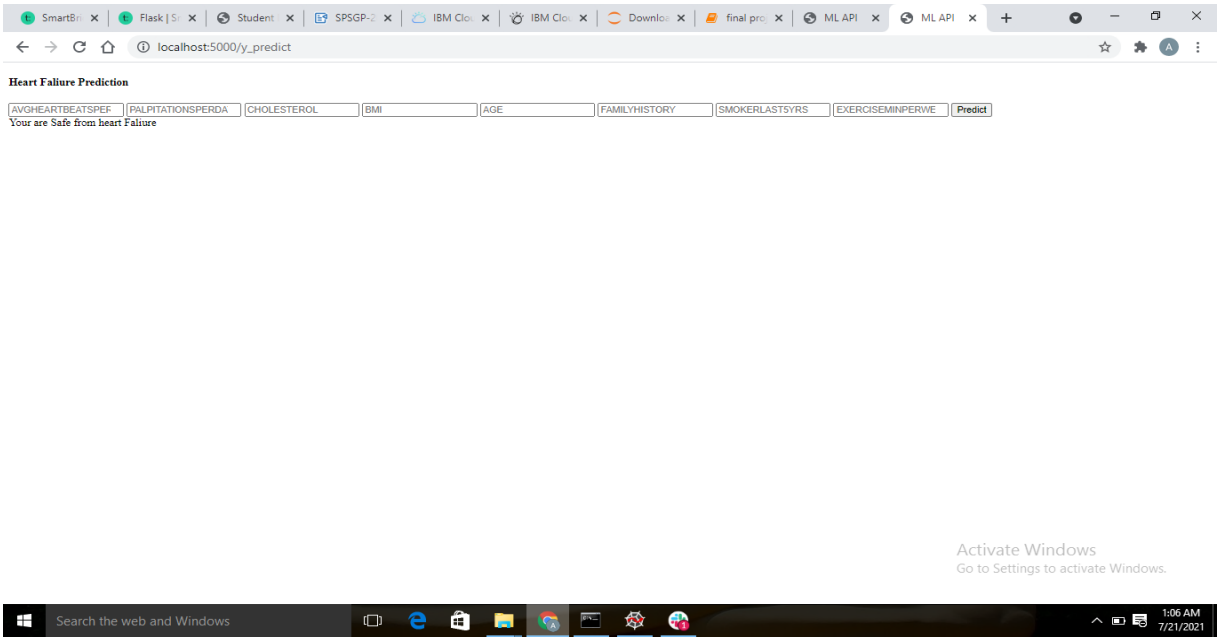
This is the project flow to complete the project

1. Log in to IBM account
2. Create IBM Watson Studio and Node-RED Service
3. Create a Watson studio project
4. ADD Auto AI Experiment
5. Run the Auto AI Experiment to build a Machine learning model on the desired dataset
6. Save the model
7. Deploy the model as a web server and generate scoring End Point.
8. build a ML model in jupyter notebook to preprocess the dataset for getting more accuracy with that dataset.
9. create a pickle file to integrate the model with flask work.
10. Create a WEB application Using flask integration to take user input and showcase Prediction on UI.

Result

- In Heart Failure Prediction, in this dataset I have used **RandomForestClassifier** for getting more accuracy and confirming the less number of false negative and false positive(confussion matrix).
- In auto AI, deployed pipeline 3 RandomForestClassifier getting 83% accuracy without balancing the dataset.
- I have developed a model in jupyter notebook with 96% accuracy with good confussion matrix.
- Here the dataset is unbalanced in nature and dataset is balanced by oversampling operations.
- groupby() is done to remove the unwanted column from the dataset. here sex column is removed to learn and train from the dataset to give good results.
- I have checked the hyperparameter tunning and default value is giving the best result.
- GridSearch CV is used to find the best permutation and combinations.
- so I decided to integrate the created ML model from jupyter notebook to app.py file by dumping my pickle file to the creating webserver.
- flask integration is used which takes inputs from the user and showcases the prediction on UI





Advantages

- AI helps in robot assisted surgeries.
- It can perform early detection of symptoms and triage.
- It can also help us in patient care and medication management.
- AI assistant can help and assist the patients by providing virtual care through online.
- AI has the capability to design proper treatment plans for the patients.
- AI can reduce human error and can able to free up the time of doctors.

Disadvantages

- Cost efficient due to increased development of machineries.
- it also affects the human interference which causes major problem in the employment.

Applications

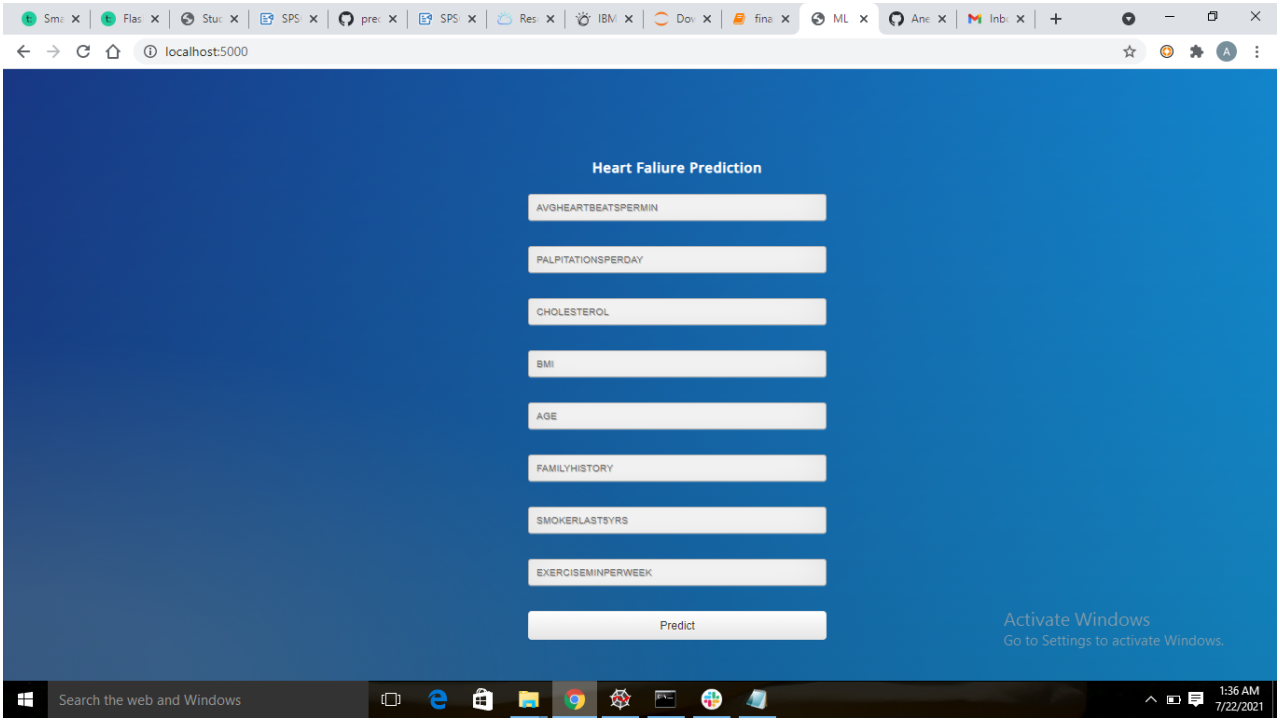
Applications of artificial intelligence in COVID-19 pandemic.Healthcare industry requires the support of new technologies like Artificial Intelligence, Internet of things etc. The main aim of AI is to analyze, prepare for the prevention measures and fight with COVID-19 (Coronavirus) and other pandemics. AI is one of the technology which can easily find the spread of the virus, identifying the high-risk patients and also useful in controlling the spread of infection in real time. The major applications of AI in COVID – 19 pandemic are early detection and diagnosis of the infection by medical imaging technologies like Computed tomography (CT), monitoring the treatment by creating an intelligent platform for automatic monitoring, prediction, neural network can also be developed to view the visual features of the disease and this would be helpful in proper monitoring and treatment of the patients, Contact tracing of the patients by analyzing the level of infection of the virus, development of drugs and vaccines by speeding up the drug testing process in real-time, where normal testing takes plenty of time which may not be possible by a human. It can be helpful in identifying the useful drugs for the treatment of COVID-19 patients. AI is not only helpful in the treatment of COVID-19 infected patients but also for the proper health monitoring and can also helpful in developing proper treatment knowledge, prevention strategies, drug and vaccine development.

Conclusion

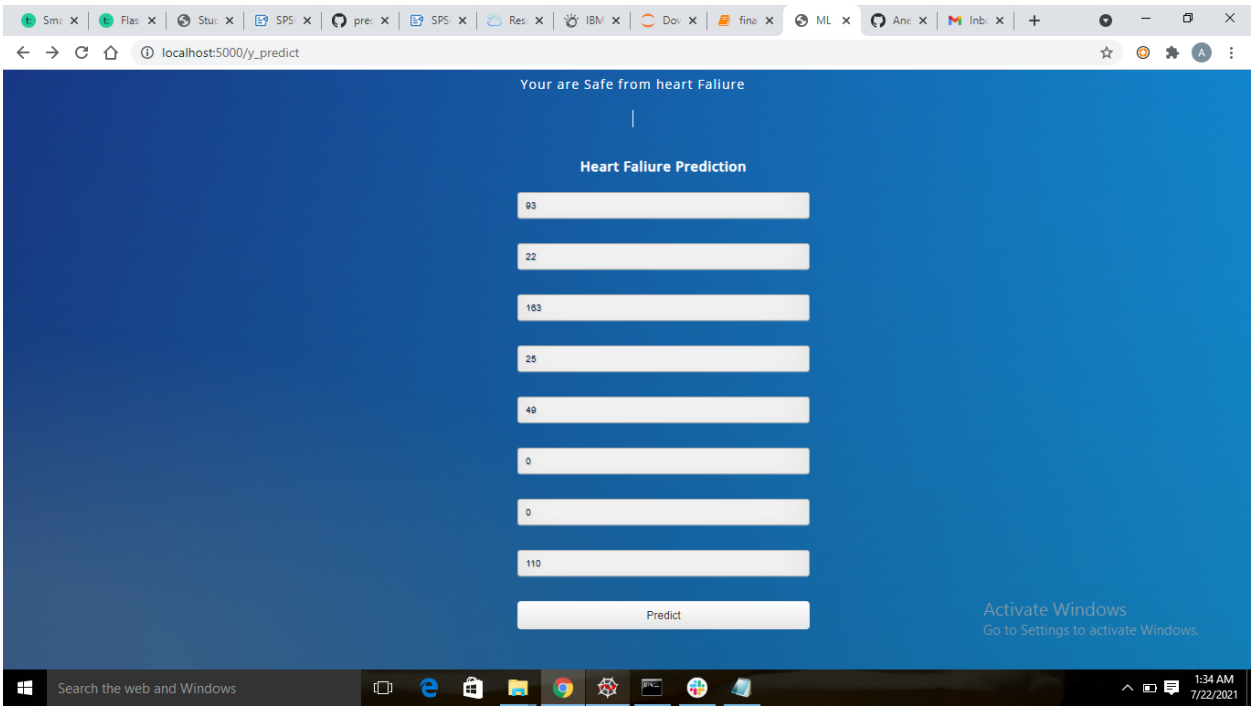
We can conclude that Artificial Intelligence is the growing and developing technology at present and has a potential to increase the development of many sectors especially in the healthcare industry. AI requires large amount of healthcare data to train and learn in order to give accurate and precise clinical decision and to increase the efficiency of the treatment. AI is used in various areas of medical field like managing healthcare records and data, drug development, treatment ways etc and also it can help the physicians to analyse and treat the patients accurately . Even though AI is used effectively, it also has some disadvantages like Cost efficient due to increased development of machineries, it also affects the human interference which causes major problem in the employment. Every new invention will have both advantages and disadvantages but we need to consider the positive thoughts to increase the development of the world. Since AI help the clinicians to take accurate and precise decision and it leads to the increase of patient care as the whole. It will be more useful for the developing

countries like India because it as more advantages for the improvement of the environment and the society . Hence, AI plays a major role in increasing the growth of healthcare industry.

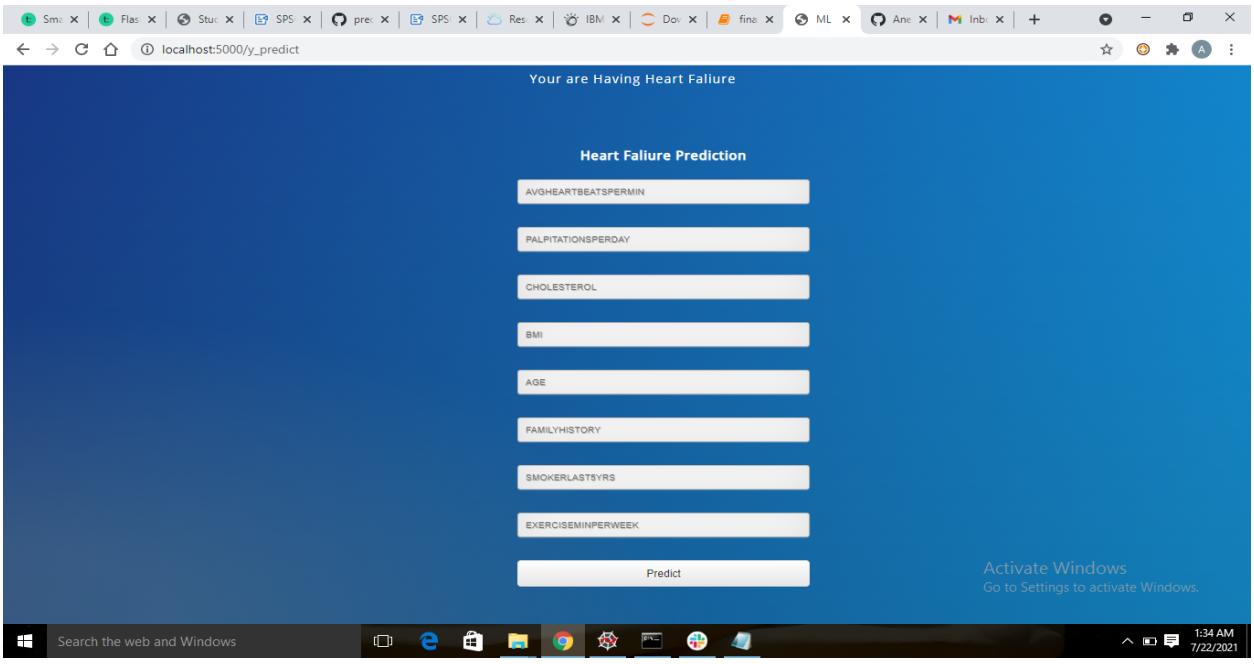
Appendix



home page of webserver



giving the inputs



output predicted

THANK YOU