Early Prediction Of Diabetes Mellitus In Pregnant Women Using IBM Auto AI

1 INTRODUCTION

1.1 Overview

Diabetes mellitus is a chronic disease characterized by hyperglycemia. It may cause many complications. According to the growing morbidity in recent years, in 2040, the world's diabetic patients will reach 642 million, which means that one of the ten adults in the future is suffering from diabetes. There is no doubt that this alarming figure needs great attention. With the rapid development of machine learning , machine learning learning has been applied to many aspects of medical health. It is a chronic disease or group of metabolic disease where a person suffers from an extended level of blood glucose in the body, which is either the insulin production is inadequate, or because the body's cells do not respond properly to insulin. The constant hyperglycemia of diabetes is related to long-haul harm, brokenness, and failure of various organs, particularly the eyes, kidneys, nerves, heart, and veins, The objective of this research is to make use of significant features. Based on the few available health parameters I can predict the women will get diabetic or non diabetic after her delivery. For this prediction I can consider the few parameters like Number of pregnancies of the women Glucose, Blood pressure, Skin thickness, Insulin, BMI, Diabetes pedigree function, Age.

1.2 Purpose

Diagnosis of diabetes is considered as a challenging problem for quantitative research. Most studies have suggested that a higher white blood cells count is due to the chronic inflammation during hypertension. A family history of diabetes has not been associated with BMI and insulin. However, an increased BMI is not always assosciated with abdominal obesity. A single parameter is not very effective to accurately diagnose diabetes and may be misleading in the decision making process. There is a need to combine different parameters to effectively predict diabetes at an early stage. Several existing techniques have not provided effective results when different parameters were used for prediction of diabetes. To review the long term effects of the diabetic pregnancy on the offspring among the Pima Indians Diabetes.

2 LITERATURE SURVEY

2.1 Existing problem

Numerous algorithms are used to predict diabetes , including the traditional machine learning method , such as support vector machine (SVM),decision tree (DT),logistic regression and so on. The random forest algorithm for the prediction of diabetes for a patient

with a higher accuracy by using random forest in machine learning technique. Nonso Nnamoko presented predicting diabetes onset an ensemble supervised learning approach they used five widely used classifiers are employed for the ensembles and a meta classifier is used to aggregate their outputs. Tejas N. Joshi presented Diabetes Prediction Using Machine Learning Techniques aims to predict diabetes via three different supervised machine learning methods including SVM, Logistic regression, ANN. This project proposes an effective technique for earlier detection of the diabetes disease.

2.2 Proposed solution

The proposed model gives the best result for diabetic prediction and the result showed that the prediction system is capable of predicting the diabetes disease effectively and efficiently. We can integrate this with Node red to make it a fully working website which can be partnered with any hospital. We have also used IBM Watson Assistant to make the feel convenient and know more about their body. The proposed method, diabetes onset prediction can be done with higher accuracy. Diabetes Prediction is becoming the area of interest for researchers in order to train the program to identify the patient are diabetic or not by applying proper classifier on the dataset. Based on previous research work, it has been observed that the classification process is not much improved. Hence a system is required as Diabetes Prediction is important area in computers, to handle the issues identified based on previous research.

3 THEORITICAL ANALYSIS

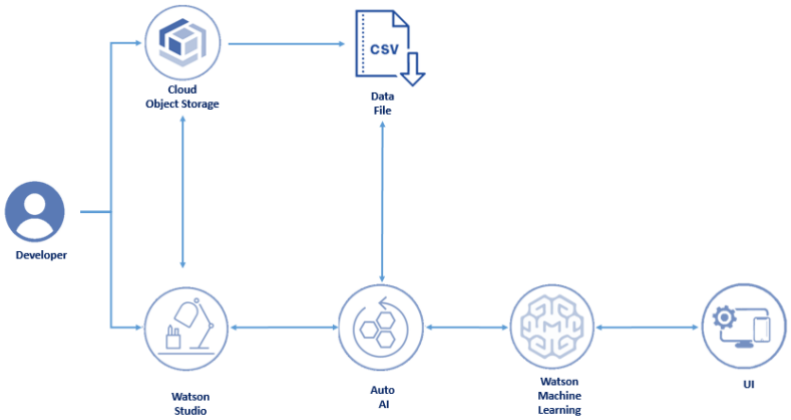
3.1 Block diagram Diagrammatic overview of the project.

3.2 Hardware / Software designing Hardware and software requirements of the project

Services Used:

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

Technical Architecture:



4 EXPERIMENTAL INVESTIGATIONS

Predict the diabetes of Pima Female Indians. The data set properties are 9 attributes representing 769 female Indians. The information that we get from the data set are several constraints were placed on the selection of these instances from a larger database, In particular, all patients here are females at least 21 years old of Pima Indian heritage.

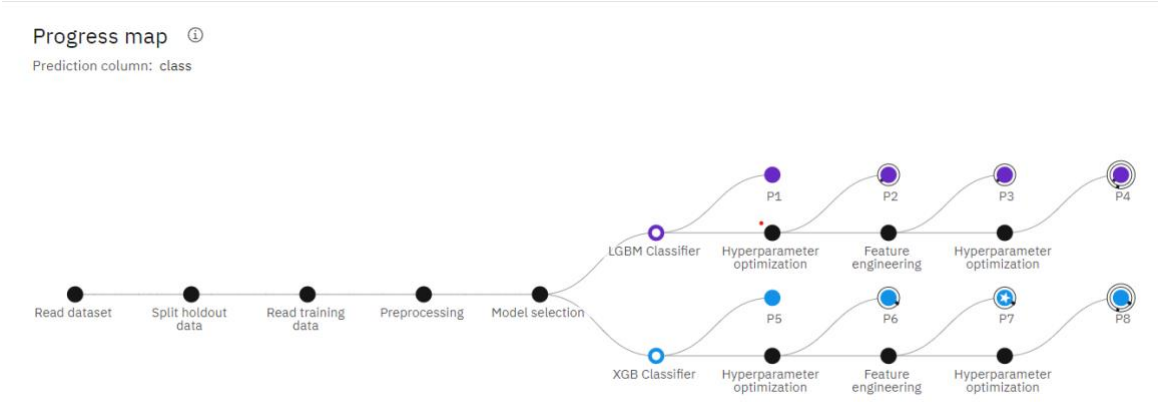
Data set Attributes are:

- Age
- BMI (column name in the dataset - mass)
- Pregnancies (column name in the dataset - preg)
- Diabetes Pedigree Function
- Glucose level (column name in the dataset - pals)
- blood pressure(column name in the dataset - pess)
- Skin Thickness(column name in the dataset - skin)
- Insulin (column name in the dataset - test)

Data set:

	A	B	C	D	E	F	G	H	I
1	preg	plas	pres	Skin	test	mass	pedi	age	class
2	6	148	72	35	0	33.6	0.627	50	1
3	1	85	66	29	0	26.6	0.351	31	0
4	8	183	64	0	0	23.3	0.672	32	1
5	1	89	66	23	94	28.1	0.167	21	0
6	0	137	40	35	168	43.1	2.288	33	1
7	5	116	74	0	0	25.6	0.201	30	0
8	3	78	50	32	88	31	0.248	26	1
9	10	115	0	0	0	35.3	0.134	29	0
10	2	197	70	45	543	30.5	0.158	53	1
11	8	125	96	0	0	0	0.232	54	1
12	4	110	92	0	0	37.6	0.191	30	0
13	10	168	74	0	0	38	0.537	34	1
14	10	139	80	0	0	27.1	1.441	57	0
15	1	189	60	23	846	30.1	0.398	59	1
16	5	166	72	19	175	25.8	0.587	51	1
17	7	100	0	0	0	30	0.484	32	1
18	0	118	84	47	230	45.8	0.551	31	1
19	7	107	74	0	0	29.6	0.254	31	1
20	1	103	30	38	83	43.3	0.183	33	0
21	1	115	70	30	96	34.6	0.529	32	1
22	3	126	88	41	235	39.3	0.704	27	0
23	8	99	84	0	0	35.4	0.388	50	0
24	7	196	90	0	0	39.8	0.451	41	1
25	9	119	80	35	0	29	0.263	29	1
26	11	143	94	33	146	36.6	0.254	51	1
27	10	125	70	26	115	31.1	0.205	41	1

pima-indians-diabetes.data

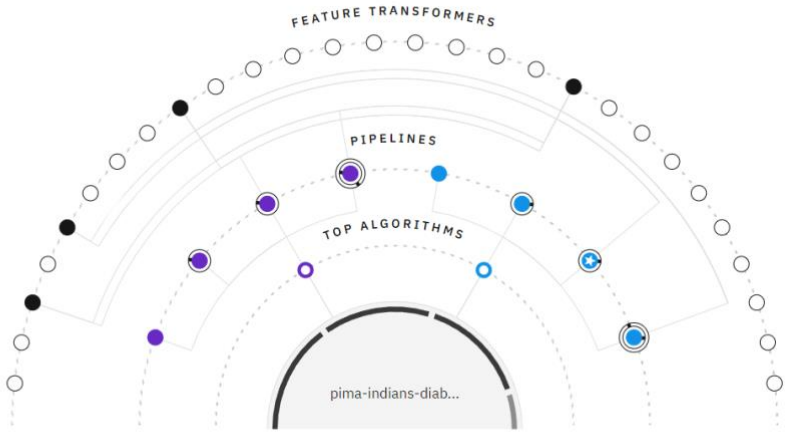


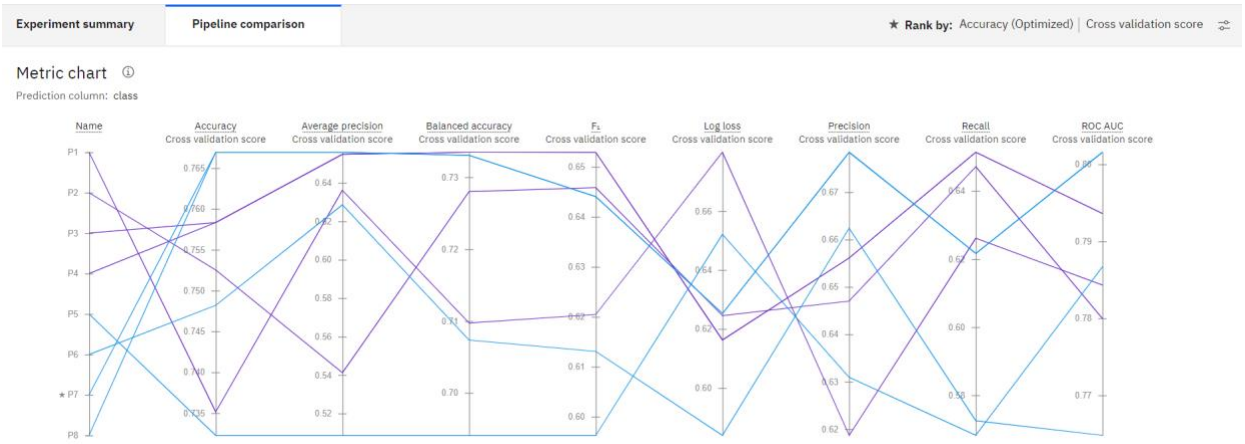
Pipeline leaderboard

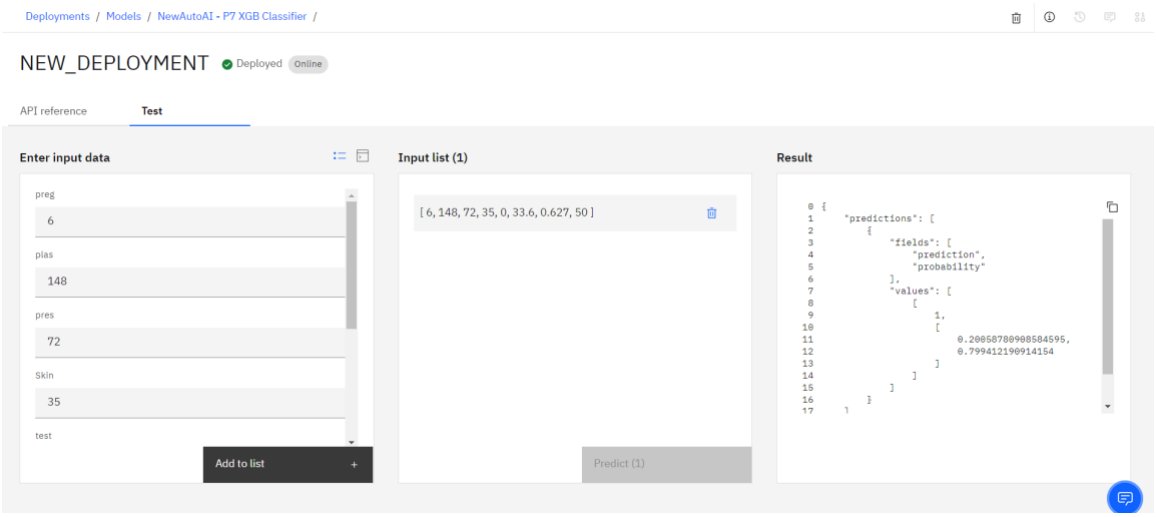
	Rank	Name	Algorithm	Accuracy (Optimized) Cross Validation	Enhancements	Build time
★	1	**Pipeline 7**	XGB Classifier	**0.767**	HPO-1 FE	00:00:38
	2	**Pipeline 8**	XGB Classifier	**0.767**	HPO-1 FE HPO-2	00:00:20
	3	**Pipeline 3**	LGBM Classifier	**0.758**	HPO-1 FE	00:00:28
	4	**Pipeline 4**	LGBM Classifier	**0.758**	HPO-1 FE HPO-2	00:00:24
	5	**Pipeline 2**	LGBM Classifier	**0.753**	HPO-1	00:00:11
	6	**Pipeline 6**	XGB Classifier	**0.748**	HPO-1	00:00:08
	7	**Pipeline 1**	LGBM Classifier	**0.735**	*None*	00:00:01

Relationship map

Prediction column: class







5 PROJECT FLOW

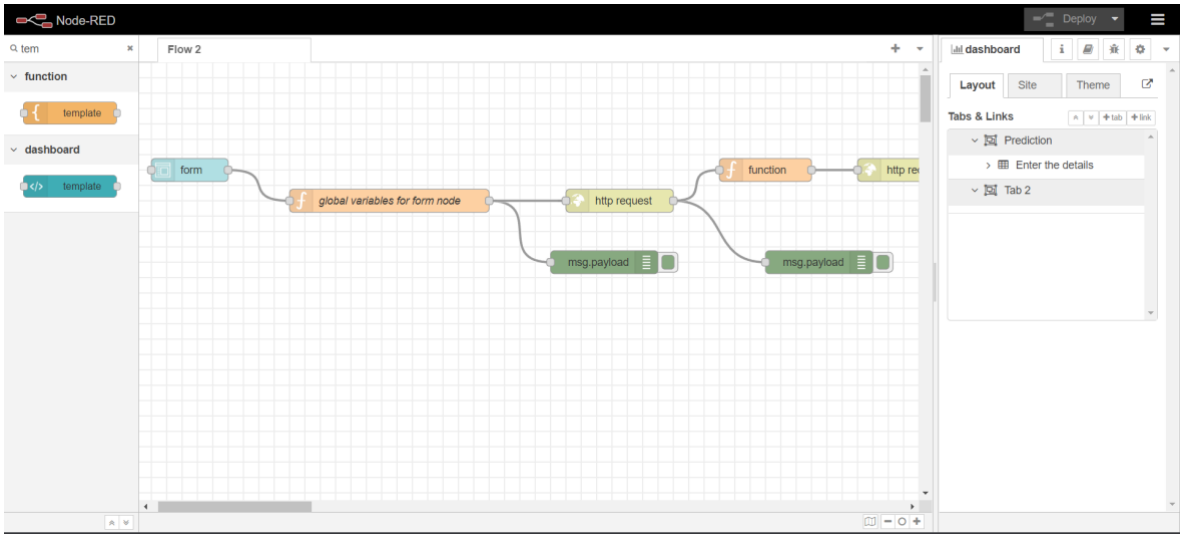
- Log in to IBM account
- Create IBM Watson Studio and Node-RED Service
- Create a Watson studio project
- ADD Auto AI Experiment
- Run the Auto AI Experiment to build a Machine learning model on the desired dataset
- Save the model

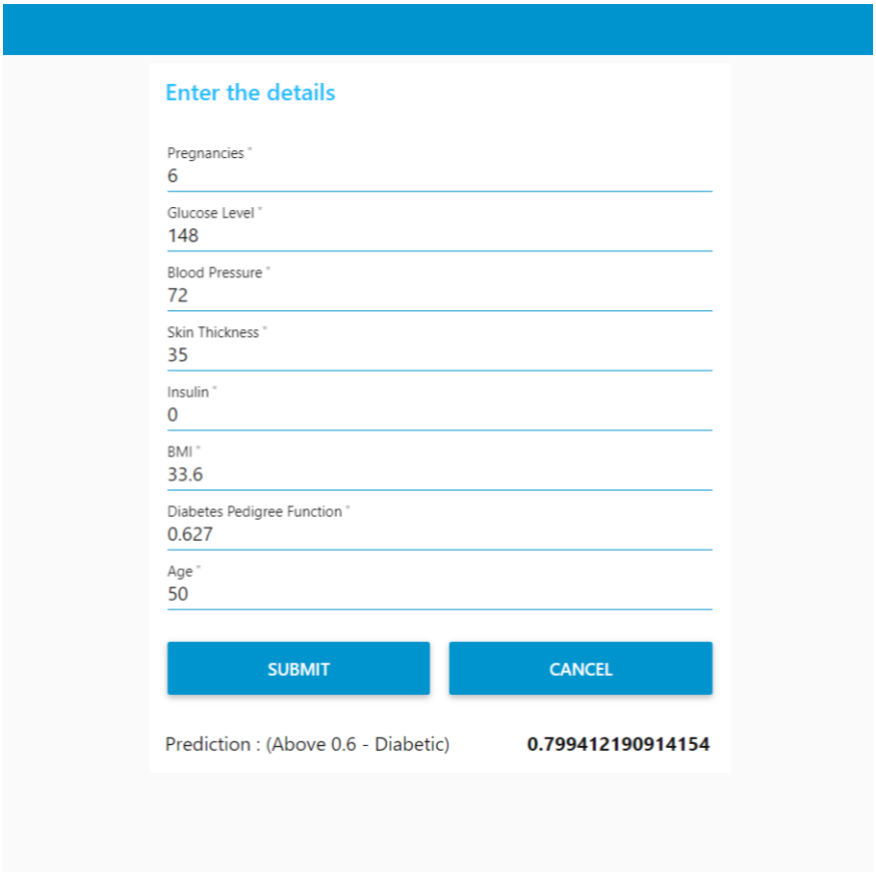
- Deploy the model as a web server and generate scoring End Point
- Create a WEB application Using Node-RED to take user input and showcase Prediction on UI

6 FLOWCHART

7 RESULT

Node RED Integration with Auto Ai Model:





8 ADVANTAGES & DISADVANTAGES

Advantages:

XGBoost is an efficient and easy to use algorithm which delivers high performance and accuracy as compared to other algorithms. XGBoost is also known as regularized version of GBM. let's see some of the advantages of the XGBoost algorithm.

Regularization: XGBoost has in-built L1(lasso regression)and L2(ridge regression)regularization which prevents the model from overfitting. That is why, XGBoost is also called regularized form of GBM(gradient Boosting Machine).
While using scikit learn library, we pass two hyper- parameters (alpha and lambda)to XGBoost related to regularization. Alpha is used for regularization and lambda is used for L2 regularization.

parallel processing : XGBoost utilizes the power of parallel processing and that is why it is much faster than GBM.it uses multiple CPU cores to execute the model.

Handling missing values : XGBoost has an in-built capability to handle missing values. When XGBoost encounters a missing values at a node, it tries both the left and right hand split and learns the way leading to higher loss for each node. It then does the same when working on the testing data.

Disadvantages:

Well XGBoost(as with other boosting techniques)is more likely to overfit than bagging does(ie; random forest)but with a robust enough dataset and conservative hyperparameters, higher accuracy is the reward. XGBoost takes quite a while to fail, that's another drawback when compared to more naive approaches .Overall though, as far as boosting goes, XGBoost is an upgrade on an idea(gradient boosting)that was itself an improvement on naive bagging techniques. Because it was created relatively recently and its design took into account the issues with existing models, it tends to outperform them based on those matrices, it's important to remember that XGBoost is essentially just regular gradient.

9 APPLICATIONS

We can integrate this with Node red to make it a fully working website which can be partnered with any hospital. We have also used IBM Watson Assistant to make the feel convenient and know more about their body.

10 CONCLUSION

The main aim of this project was to design and implement Diabetes Prediction using Machine Learning methods and performed analysis of that methods and it has been achieved successfully. The proposed approach uses XGBoost classifier with hyperparameter tuning and feature engineering using IBM Auto AI service. And 77% classification accuracy has been achieved. The experimental results can be assist health care to take early prediction and make early decision to cure diabetes and save humans life.

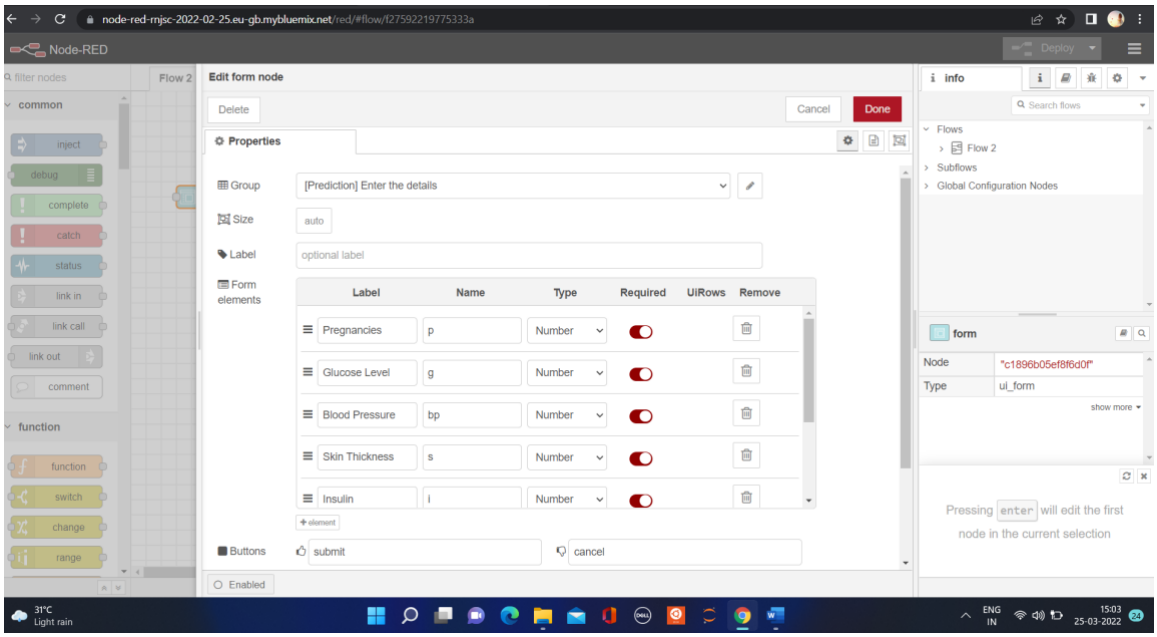
11 FUTURE SCOPE

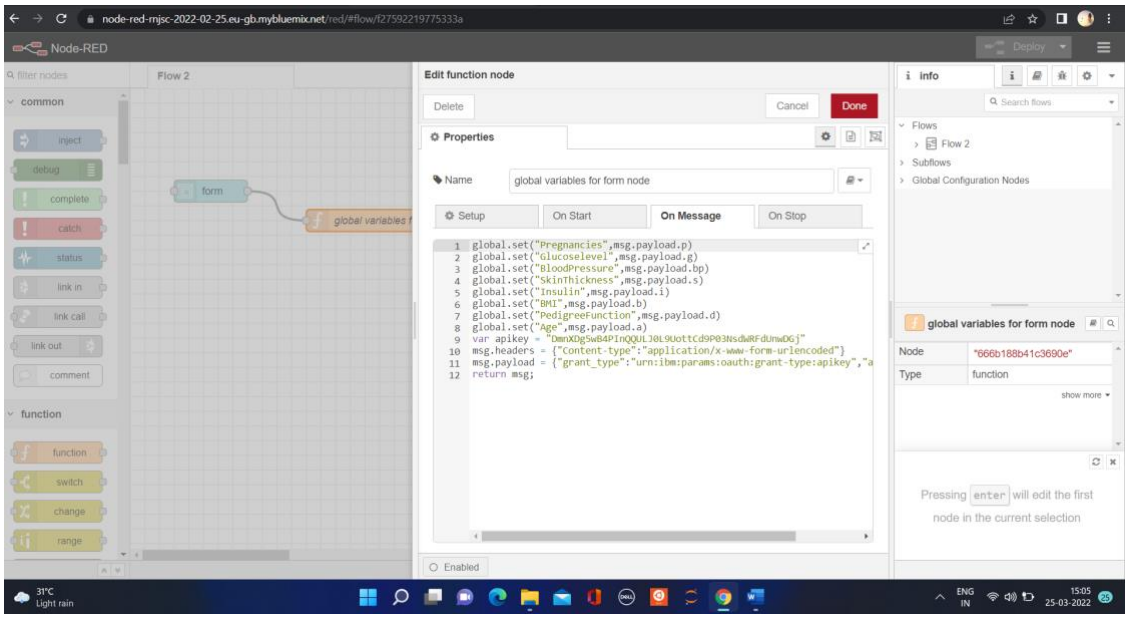
In future if we get a large set of diabetic dataset we can perform comparative analysis of analysing the performance of each algorithm as well as the hybrid algorithm so that the best one can be applied for predictive analysis. A particular method to identify diabetes is not very sophisticated way for initial diabetes detection and it is not fully accurate for predicting diseases. That’s why we need a smart hybrid predictive analytics diabetes diagnostic system that can effectively work with accuracy and efficiency. We can use datamining, neural networks for exploring and utilizing to support medical decision, which improves in diagnosing the risk for pregnant diabetes. Due to the dataset we have till date are not up to the mark we cannot predict the type of diabetes, so in future we aim to predicting type of diabetes and explore it, which may improve the accuracy of predicting diabetes. We can also study the causes of diabetes and how to avoid having diabetes.

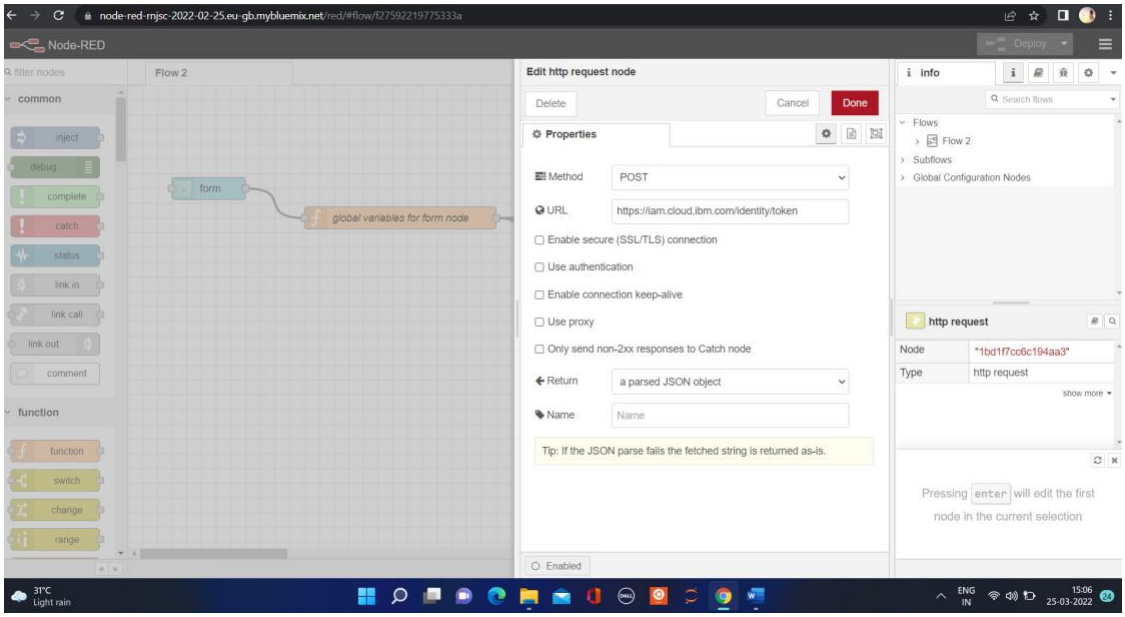
12 BIBILOGRAPHY

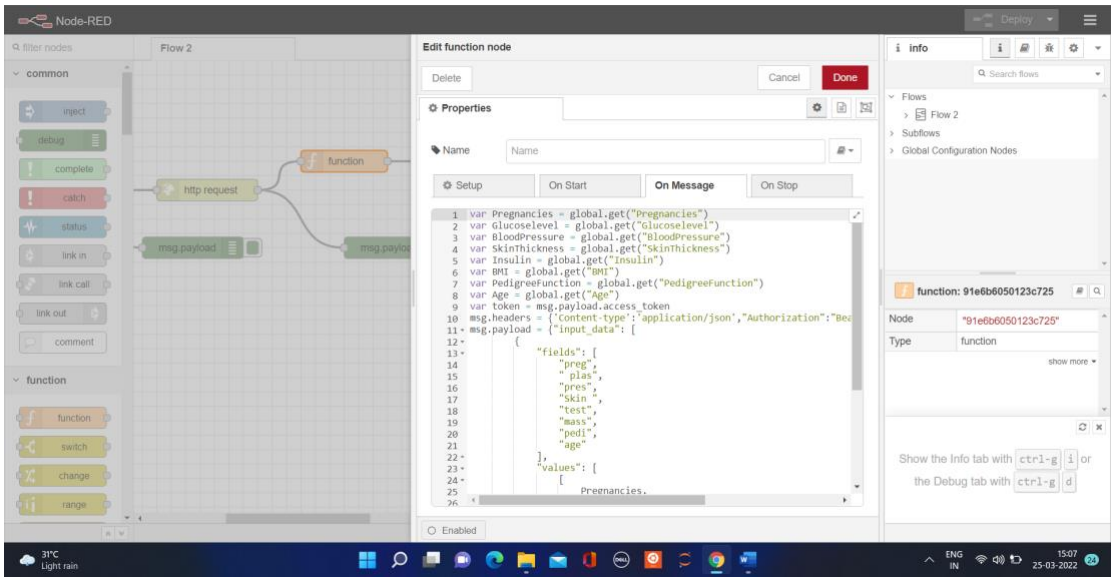
- Nonso Nnamoko, Abir Hussain, David England, "Predicting Diabetes Onset: an EnsembleSupervised Learning Approach IEEE Congress on Evolutionary Computation (CEC), 2018.
- Tejas N. Joshi, Prof. Pramila M. Chawan, "Diabetes Prediction Using Machine LearningTechniques".Int. Journal of Engineering Research and Application, Vol. 8, Issue 1. (Part-II) January 2018, pp.-09-13

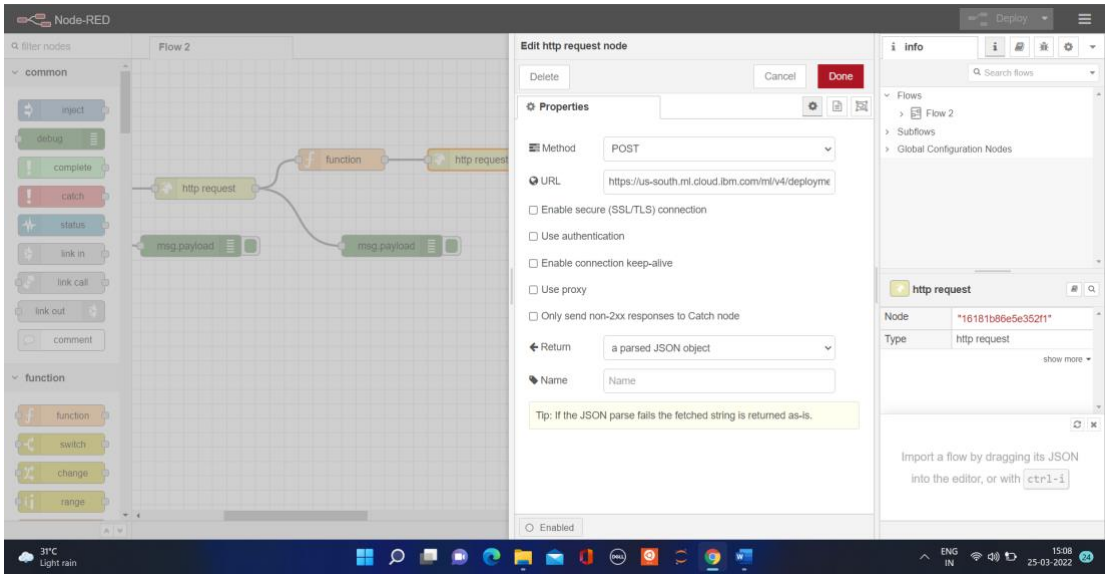
APPENDIX

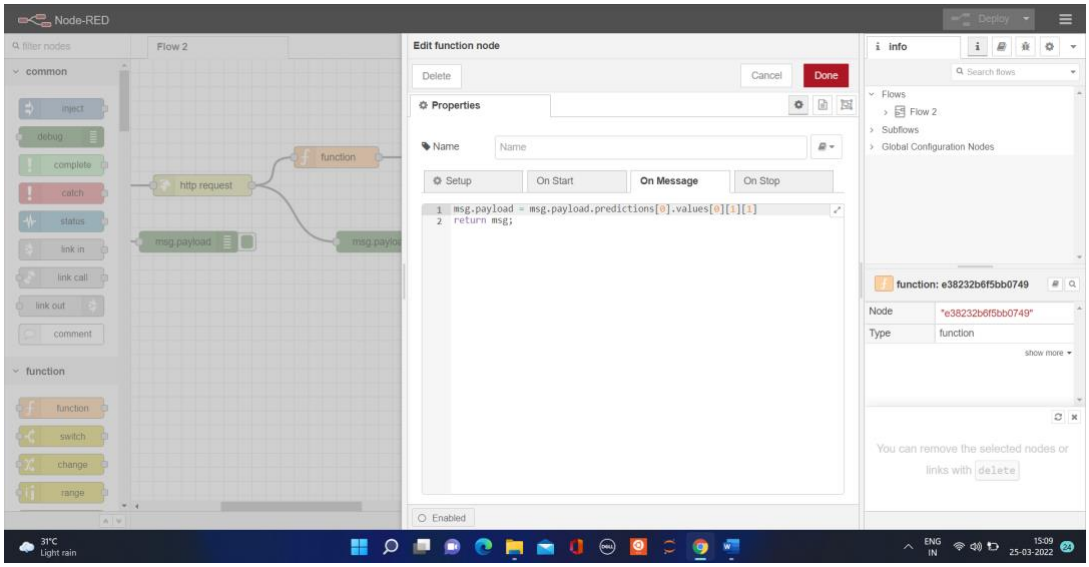


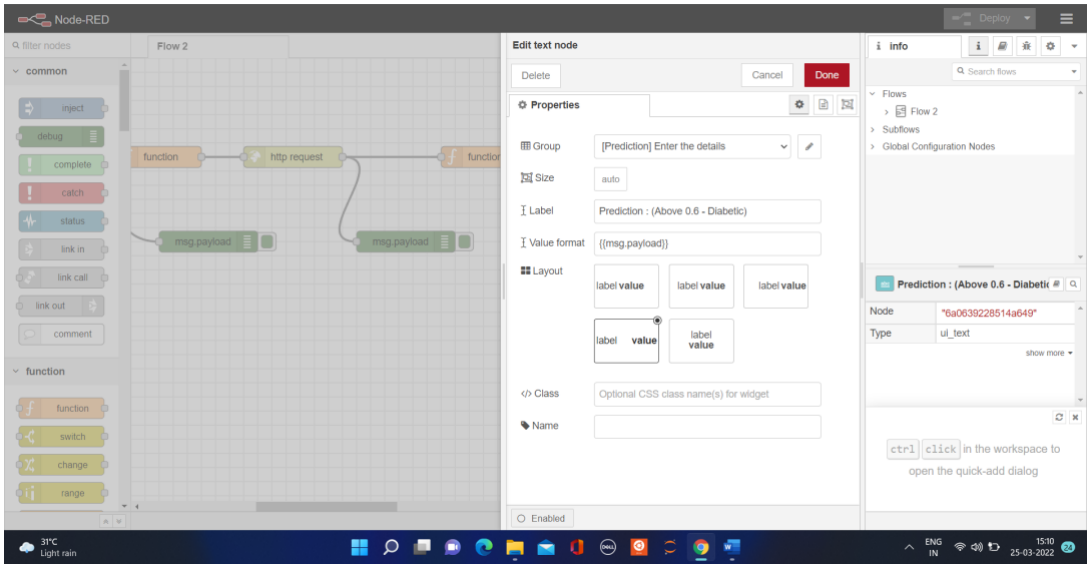



node-red-rnjsc-2022-02-25.eu-gb.mybluemix.net/red/#flow/f27592219775333a
Node-RED
Deploy
filter nodes
Flow 2
common
inject
debug
complete
catch
status
link in
link call
link out
comment
function
function
switch
change
range
form
global variables for form node
Edit http request node
Delete
Cancel
Done
Properties
Method
POST
URL
https://iam.cloud.ibm.com/identity/token
Enable secure (SSL/TLS) connection
Use authentication
Enable connection keep-alive
Use proxy
Only send non-2xx responses to Catch node
Return
a parsed JSON object
Name
Name
Tip: If the JSON parse fails the fetched string is returned as-is.
Enabled
info
Search flows
Flows
Flow 2
Subflows
Global Configuration Nodes
http request
Node
"1bd1f7cc6c194aa3"
Type
http request
show more
Pressing enter will edit the first node in the current selection
31°C
Light rain
ENG
IN
15:06
25-03-2022


Node-RED
Deploy
filter nodes
Flow 2
common
inject
debug
complete
catch
status
link in
link call
link out
comment
function
function
switch
change
range
http request
function
msg.payload
Edit function node
Delete
Cancel
Done
Properties
Name
Name
Setup
On Start
On Message
On Stop
var Pregnancies = global.get("Pregnancies")
var Glucoselevel = global.get("Glucoselevel")
var BloodPressure = global.get("BloodPressure")
var SkinThickness = global.get("SkinThickness")
var Insulin = global.get("Insulin")
var BMI = global.get("BMI")
var PedigreeFunction = global.get("PedigreeFunction")
var Age = global.get("Age")
var token = msg.payload.access_token
msg.headers = {'content-type':'application/json','Authorization":"Bea
msg.payload = {"input_data": [
{
"fields": [
"preg",
" plas",
"pres",
"skin ",
"test",
"mass",
"pedi",
"age"
],
"values": [
[
Pregnancies,
Enabled
info
Search flows
Flows
Flow 2
Subflows
Global Configuration Nodes
function: 91e6b6050123c725
Node
"91e6b6050123c725"
Type
function
show more
Show the Info tab with ctrl-g i or the Debug tab with ctrl-g d
31°C
Light rain
ENG
IN
15:07
25-03-2022


Node-RED
Deploy
Flow 2
common
inject
debug
complete
catch
status
link in
link call
link out
comment
function
switch
change
range
http request
function
msg.payload
Edit http request node
Delete
Cancel
Done
Properties
Method
POST
URL
https://us-south.ml.cloud.ibm.com/ml/v4/deployme
Enable secure (SSL/TLS) connection
Use authentication
Enable connection keep-alive
Use proxy
Only send non-2xx responses to Catch node
Return
a parsed JSON object
Name
Tip: If the JSON parse fails the fetched string is returned as-is.
Enabled
info
Search flows
Flows
Flow 2
Subflows
Global Configuration Nodes
http request
Node
"16181b86e5e352f1"
Type
http request
show more
Import a flow by dragging its JSON into the editor, or with ctrl-i
31°C
Light rain
ENG
IN
15:08
25-03-2022


Node-RED
Deploy
Flow 2
common
function
switch
change
range
http request
Edit function node
Delete
Cancel
Done
Properties
Name
Setup
On Start
On Message
On Stop
1 msg.payload = msg.payload.predictions[0].values[0][1][1]
2 return msg;
Enabled
info
Flows
Flow 2
Subflows
Global Configuration Nodes
function: e38232b6f5bb0749
Node
"e38232b6f5bb0749"
Type
function
show more
You can remove the selected nodes or links with delete
31°C
Light rain
ENG
IN
15:09
25-03-2022


Node-RED
Deploy
filter nodes
Flow 2
common
inject
debug
complete
catch
status
link in
link call
link out
comment
function
switch
change
range
function
http request
msg.payload
msg.payload
Edit text node
Delete
Cancel
Done
Properties
Group
[Prediction] Enter the details
Size
auto
Label
Prediction : (Above 0.6 - Diabetic)
Value format
{{msg.payload}}
Layout
label value
label value
label value
label value
label value
Class
Optional CSS class name(s) for widget
Name
Enabled
info
Search flows
Flows
Flow 2
Subflows
Global Configuration Nodes
Prediction : (Above 0.6 - Diabeti
Node
"6a0639228514a649"
Type
ui_text
show more
ctrl click in the workspace to open the quick-add dialog
31°C
Light rain
ENG
IN
15:10
25-03-2022