

SHESHADRI RAO GUDLAVALLERU ENGINEERING COLLEGE



EARLIER PREDICTION OF DIABETES MELLITUS IN PREGNANT WOMEN USING IBM AUTO AI



SUBMITTED BY:

G.Kushwanth kumar

ROLL NO: 19481A0333

CONTENT

1.INTRODUCTION

2.LITERATURE SURVEY

3.THERIOTICAL ANALYSIS

4.PROPOSED METHOD

5.EXPERIMENTAL INVESTIGATIONS

6.PROJECT FLOW

7.RESULT

8.ADVANTAGES AND DISADVANTAGES

9.APPLIACATIONS

10.CONCLUSION

11.FUTURE SCOPE

1.INTRODUCTION

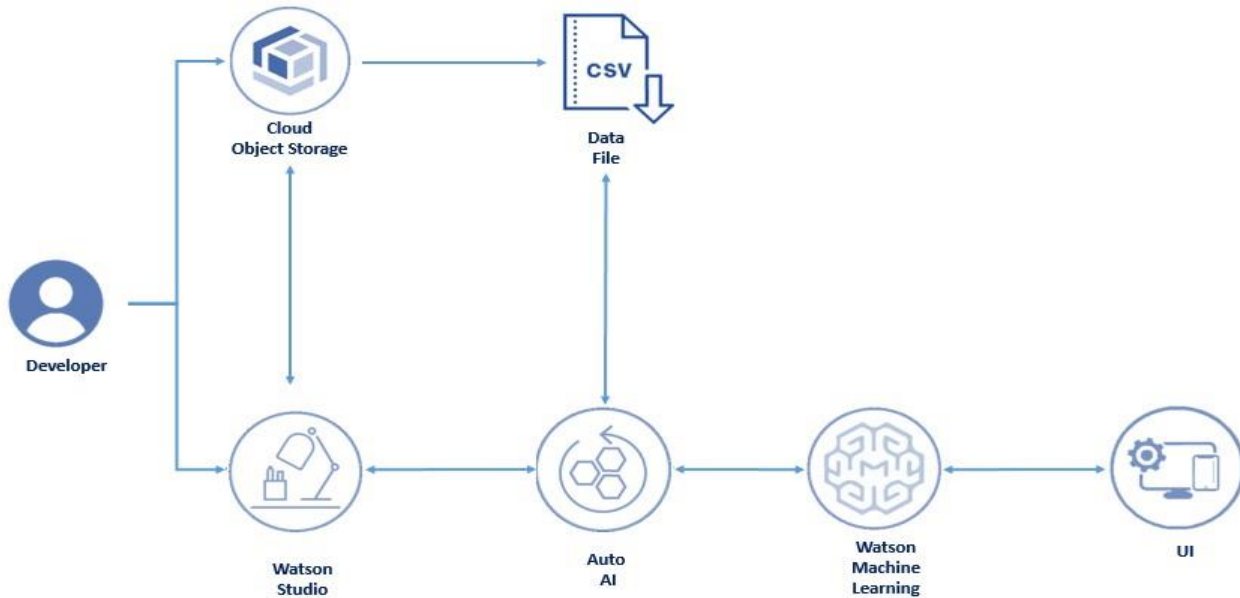
2.LITERATURE SURVEY

2.1 Existing problem

Recently, 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 some people proposed random forest algorithm for the prediction of diabetes develop a system which can perform early prediction of diabetes for a patient with higher accuracy by using random forest algorithm in machine learning technique. 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, efficiently and most importantly instantly. Nonso Nnamoko et al. presented predicting diabetes onset: an ensemble supervised learning approach they used five widely used classifiers are employed for the ensembles and a metaclassifiers is used to aggregate their outputs. The results are presented and compared with similar studies that used the same dataset with in the literature .It is shown that by a proposed method, diabetes onset prediction can be done with higher accuracy .Deeraj Shetty proposed diabetes disease prediction using data mining assemble intelligent Diabetes disease prediction system that gives analysis of diabetes malady utilizing diabetes patients database in this system they proposed the use of algorithms like Bayesian and KNN(K nearest neighbor to apply on diabetes patients database and analyze them by taking various attributes of diabetes for prediction of diabetes disease. Comparison of different machine learning techniques used in the study reveals which algorithm is best suited for prediction of diabetes. 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 s system is required as Diabetes Prediction is important area in computers, to handle the issue identified on previous research.

3.THERIOTICAL ANALYSIS

3.1 Technical Architecture



3.2 Hardware/Software designing services used

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

4.PROPOSED METHOD

4.1 dataset

Student Dashboard

Service Details - IBM Cloud

IBM Watson Studio

← → ↻

dataplatfom.cloud.ibm.com/projects/6be2936d-2b8e-4df8-90c9-2cdbae95fc27/data-assets/028aa05e-c77a-4591-9491-6926bd477df7/preview?context=cpdaas&walkme_guided_tutorial=false

🔗 ☆ 🗑️ 🧑

IBM Watson Studio

🔍 Search in your workspaces

Buy 🔄 ⚙️ 🔔

GUDIMETLA NAGA HEMAN... ▾

Dallas ▾

GN

My projects / early prediction of Diabetes Mell... / pima-indians-diabetes.data.csv

🔍 🗑️ ⌚ 💬 🧑

Preview

Visualization

Schema: 9 Columns

The preview includes only a limited set of columns and rows. ⓘ

Last refresh: 10 seconds ago 🔄

Prepare data

preg String	plas String	pres String	Skin String	test String	mass String	pedi String	age String	class String
6	148	72	35	0	33.6	0.627	50	1
1	85	66	29	0	26.6	0.351	31	0
8	183	64	0	0	23.3	0.672	32	1
1	89	66	23	94	28.1	0.167	21	0
0	137	40	35	168	43.1	2.288	33	1
5	116	74	0	0	25.6	0.201	30	0
3	78	50	32	88	31	0.248	26	1
10	115	0	0	0	35.3	0.134	29	0
2	197	70	45	543	30.5	0.158	53	1
8	125	96	0	0	0	0.232	54	1
4	110	92	0	0	37.6	0.191	30	0
10	168	74	0	0	38	0.537	34	1
10	139	80	0	0	27.1	1.441	57	0
1	180	60	22	846	20.1	0.208	50	1

Information

×

Data asset

pima-indians-diabetes.data.csv

Description

No description available for this asset

Tags

No tags available for this asset

Creator

GUDIMETLA NAGA HEMANTH KUMAR

Usage

Created on Jul 25, 2022, 07:42 PM

Size

23.056 KB

87°F Mostly cloudy

🪟 🔍 🗑️ 📧 📁 🌐 📧 📧

ENG US 📶 🔊 🔌 3:35 PM 8/7/2022 🌙

No of times pregnant, plasma glucose concentration a 2 hours in an oral glucose tolerance test, Diastolic blood pressure(mm Hg), Triceps skin fold thickness(mm), 2-hour serum insulin(mu U/ml), Body mass index(weight in kg /(height in m)^2),Diabetes pedigree function, Age(years), class variable(0 or 1)

4.2 IBM Watson Studio

The screenshot displays the IBM Watson Studio web interface in a browser window. The browser's address bar shows the URL: `dataplatfrom.cloud.ibm.com/home2?context=cpdaas&apps=data_science_experience&nocache=true&quick_start_target=data_science_experience`. The user is logged in as GUDIMETLA NAGA HEMA.

Welcome, Gudimetla!

Take a tutorial
Step through implementing a Data fabric use case in a sample project.
→

Work with data
Create a project for your team to prepare data, find insights, or build models.
→

Learn what's new
Stay current with new features, enhancements, and other changes.
→

Quick start

- Build dashboards with IBM Cognos Dashboard Embedded
- Create data pipelines with DataStage
- Build customer profiles with IBM Match 360 with Watson
- Catalog and govern data with Watson Knowledge Catalog
- Build and manage ML models with Watson Studio

Projects

Project Name	Last Updated
early prediction of Diabetes Mellitus In Pregnent Women	Jul 25, 2022 07:33 PM

New in gallery

SAMPLE PROJECT

Text Analysis with Watson Natural Language...

Notifications

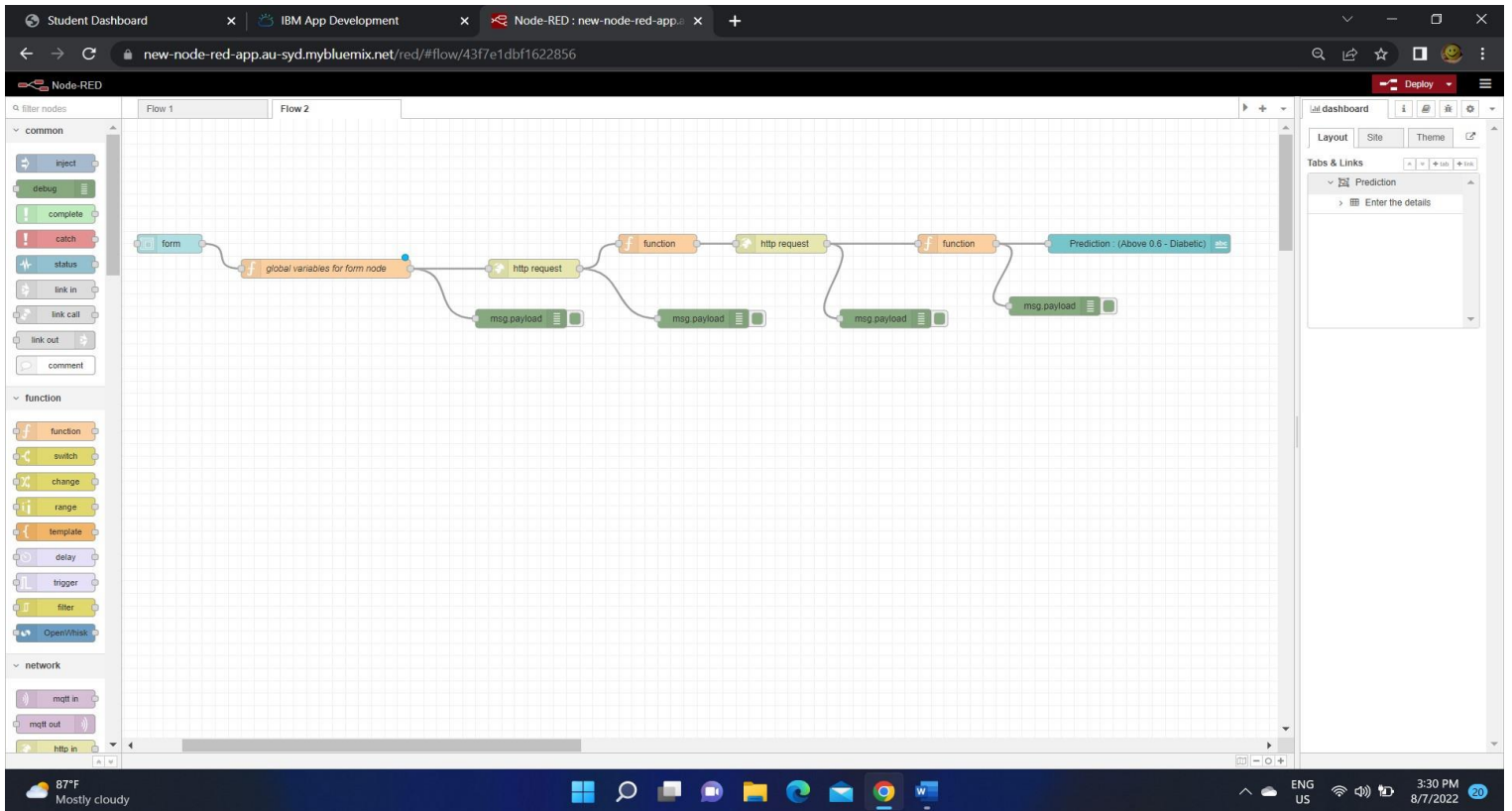
- Online deployment ready: The online deployment NEW DEPLOYMENT in space modal1 is ready to accept requests. Aug 02, 2022 06:10 PM
- Online deployment ready: The online deployment NEW DEPLOYMENT in space modal1 is ready to accept requests. Aug 02, 2022 06:10 PM
- Online deployment ready: The online deployment NEW DEPLOYMENT in space modal1 is ready to accept requests. Aug 02, 2022 06:10 PM

Deployments

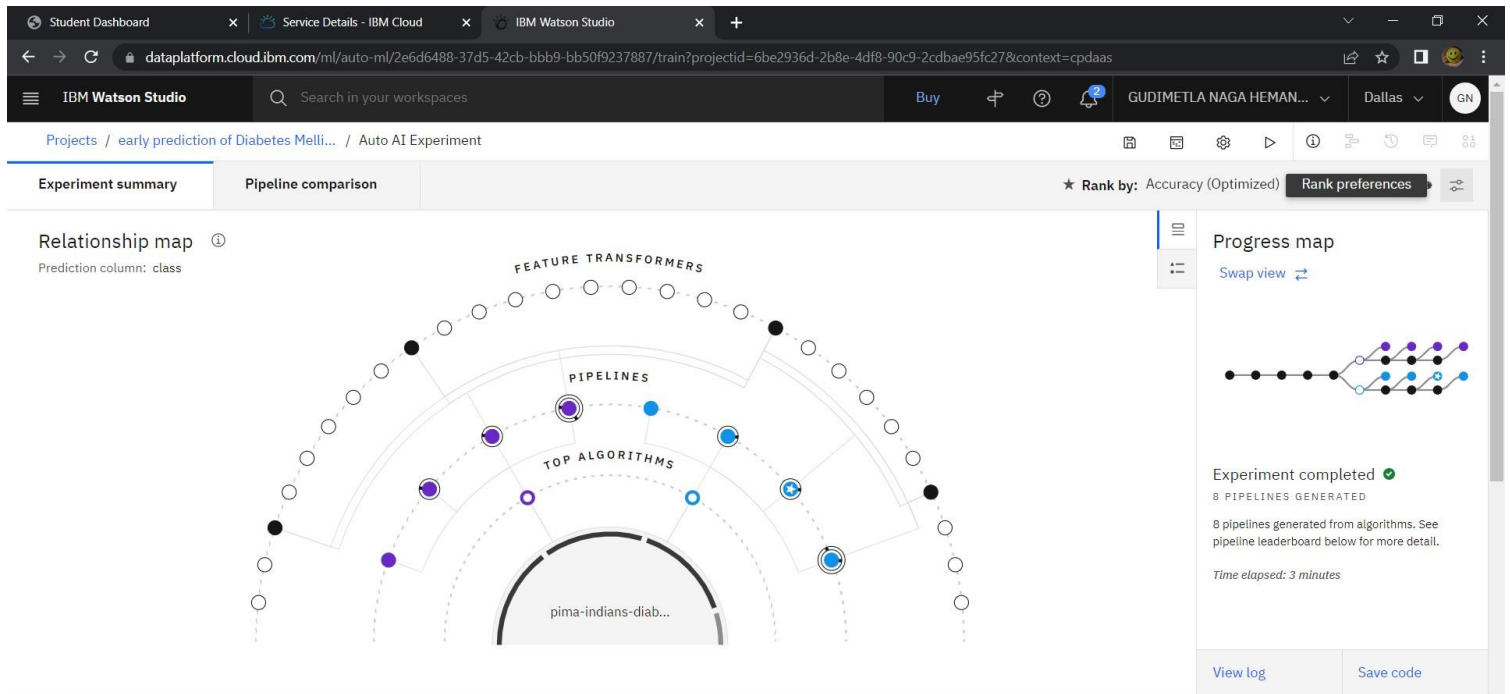
Deployment Name	Last Updated
modal1	Jul 28, 2022 01:02 PM

The bottom of the image shows a Windows taskbar with the system clock at 3:24 PM on 8/7/2022, and weather information indicating 87°F and mostly cloudy conditions.

4.2 Node-RED



5.EXPERIMENTAL INVESTIGATIONS



Pipeline leaderboard ▾

87°F Mostly cloudy | 3:28 PM 8/7/2022

Student Dashboard | Service Details - IBM Cloud | IBM Watson Studio

dataplatform.cloud.ibm.com/ml/2e6d6488-37d5-42cb-bbb9-bb50f9237887/train?projectId=6be2936d-2b8e-4df8-90c9-2cdbae95fc278&context=cpdaas

IBM Watson Studio | Search in your workspaces | Buy | ? | GUDIMETLA NAGA HEMAN... | Dallas | GN

Projects / early prediction of Diabetes Melli... / Auto AI Experiment

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

	Rank	Name	Algorithm	Accuracy (Optimized) Cross Validation	Enhancements	Build time
★	1	Pipeline 7	Snap Boosting Machine Classifier	0.763	HPO-1 FE	00:00:32
	2	Pipeline 8	Snap Boosting Machine Classifier	0.763	HPO-1 FE HPO-2	00:00:42
	3	Pipeline 3	LGBM Classifier	0.758	HPO-1 FE	00:00:43
	4	Pipeline 4	LGBM Classifier	0.758	HPO-1 FE HPO-2	00:01:09
	5	Pipeline 2	LGBM Classifier	0.753	HPO-1	00:00:11
	6	Pipeline 6	Snap Boosting Machine Classifier	0.748	HPO-1	00:00:04
	7	Pipeline 1	LGBM Classifier	0.735	None	00:00:01
	8	Pipeline 5	Snap Boosting Machine Classifier	0.734	None	00:00:01

87°F Mostly cloudy | 3:28 PM 8/7/2022

We tried different algorithms like Snap Boosting Machine Classifier, LightGBM Classifier, Decision Tree, Logistic Regression. We used 8 Pipelines to compare the performance of each model. Here 85% of the data is used for training and rest 15% is used for testing which yielded the best performance. Among all the models Snap Boosting Machine Classifier gave the best accuracy score of 76.3%

6.PROJECT FLOW

1. Login to IBM Cloud Account
2. Create IBM Watson Studio and Node-RED Services
3. Create Watson Studio Project
4. ADD auto AI Experiment
5. Run 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. Create a WEB-Application using Node-RED to take user input and showcase Prediction on UI

7.RESULT

Prediction

Enter the details

Pregnancies *

Glucose Level *

Blood Pressure *

Skin Thickness *

Insulin *

BMI *

Diabetes Pedigree Function *

Age *

Prediction : (Above 0.6 - Diabetic) 0.7549677859449343

8.ADVANTAGES AND DISADVANTAGES

8.1 Advantages

1. simple and easy to implement
2. testing is very fast
3. less over fitting in practice
4. it is more effective
5. it delivers high performance and accuracy as compared to the other algorithms
6. flexibility in the choice of weak learners, Boosting scheme
7. it can be applied to wide range of problems
8. It is flexible -it can be combined with any learning algorithm
9. It has excellent predictive accuracy
- 10.Harder to overfit
- 11.Efficient handling of data.....etc

8.2 Disadvantages

1. Needs many training examples
2. Learning is slow
3. It is sensitive to outliers
4. It requires more data for testing and training for accurate prediction
5. Sensitive to overfitting if the data is noisy

9.APPLICATIONS

We can integrate this with Node-RED to make it fully working website which can be partnered with any hospital. We have can also use Watson Assistant to make a chatbot to interact with the users and make the feel convenient and know more about their body

10.CONCLUSION

The main Aim of the project is to implement earlier prediction of diabetes mellitus in pregnant women Using Machine Learning methods and Performance Analysis of that methods and it has been Achieved Successfully. The Proposed approach uses Snap Boosting machine Classifier Using IBM Auto AI Services. And 76.3% classification accuracy has been achieved. The Experimental Results can be asst health care to take earlier prediction and make early decisions to cure diabetes and save Human life.

11.FUTURE SCOPE

In future, if we get large set of diabetic data set we can perform comparative analysis the performance of each algorithm as well as Hybrid algorithm so that the best one can applied for Predictive analysis. A particular method to identify diabetics 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 network for exploring and utilizing to support medical decision, which improves in diagnosing the risk for pregnant diabetics. Due to the data set we have till the date are not up to the mark, we cannot predict the type of diabetes, so in future we aimed to predicting type of diabetes and explore it, which may improve the accuracy of the accuracy of predicting the diabetes. We can also study the causes of diabetes and hoe to avoid having diabetics.

12.BIBLIOGRAPHY

Installation of Anaconda Navigator:

<https://www.youtube.com/embed/5mDYijMfSzs> **data collection**

<https://www.kaggle.com/datasets/akhilalexander/diabeticprediction>

IBM Watson studio service creation

<https://www.youtube.com/watch?v=aAT3qALIQXM&t=10s>

creation of Node-RED services

<https://www.youtube.com/watch?v=beCCPIH0-8c>

Create Watson studio project

<https://www.youtube.com/watch?v=Tr82IMw7BZw>

Creation of AUTO AI project

<https://www.youtube.com/watch?v=QhLNCXj5PY4>

Run Auto AI Experiment

<https://www.youtube.com/watch?v=0nKGHGJcvls>

Save the model <https://www.youtube.com/watch?v=xkH7fWItV0>

Model Deployment <https://www.youtube.com/watch?v=mqeH8-zl24I>

Basics of Node-RED

<https://www.youtube.com/watch?v=29qvSy7evgY>

Node-RED integration with Auto AI model

https://www.youtube.com/watch?v=7e6Z99-5e_E