



Diabetic Mellitus Prediction using IBM AutoAI

A

Internship Project Report

By

NINEL BERT COSTA J

RSIP Career Basic ML 087

Project ID :SPS_PRO_289

**RAMCO INSTITUTE OF TECHNOLOGY
ELECTRONICS AND COMMUNICATION ENGINEERING
RAJAPALAYAM ,TAMILNADU**

JUNE 2020

INDEX

1. ACKNOWLEDGMENT
2. INTRODUCTION
3. LITERATURE SURVEY
4. EXPERIMENTAL INVESTIGATION
5. FLOW CHART
6. RESULT
7. ADVANTAGES & DISADVANTAGES
8. APPLICATION
9. CONCLUSION
10. FUTURE SCOPE
11. JSON FILE
12. NODE RED WEB APPLICATION LINK

13. REFERENCES

ACKNOWLEDGMENT

The internship opportunity I had with SMART INTERNZ was a great chance for learning and professional development.

Therefore, I consider myself as a very lucky individual as I was provided with an opportunity to be a part of it. I am also grateful for having a chance to acquire knowledge from many wonderful people and professionals who led me through this internship period.

Bearing in mind previous I am using this opportunity to express my deepest gratitude and special thanks to the this platform in spite of being extraordinarily busy with their duties, took time out to hear, guide and keep me on the correct path and allowing me to carry out my project at their esteemed organization and extending during the training.

I express my deepest thanks to Mr.Ram Mohan Bethi, for taking part in useful decision & giving necessary advices and guidance and arranged all facilities to make life easier. I choose this

moment to acknowledge his contribution gratefully.

INTRODUCTION

Diabetes is a common chronic disease and poses a great threat to human health. The characteristic of diabetes is that the blood glucose is higher than the normal level, which is caused by defective insulin secretion or its impaired biological effects, or both. Diabetes can lead to chronic damage and dysfunction of various tissues, especially eyes, kidneys, heart, blood vessels and nerves. Diabetes can be divided into two categories, type 1 diabetes (T1D) and type 2 diabetes (T2D). Patients with type 1 diabetes are normally younger, mostly less than 30 years old. The typical clinical symptoms are increased thirst and frequent urination, high blood glucose levels. This type of diabetes cannot be cured effectively with oral medications alone and the patients are required insulin therapy. Type 2 diabetes occurs

more commonly in middle-aged and elderly people, which is often associated with the occurrence of obesity, hypertension, dyslipidemia, arteriosclerosis, and other diseases .

With the development of living standards, diabetes is increasingly common in people's daily life. Therefore, how to quickly and accurately diagnose and analyze diabetes is a topic worthy studying. In medicine, the diagnosis of diabetes is according to fasting blood glucose, glucose tolerance, and random blood glucose levels. The earlier diagnosis is obtained, the much easier we can control it. Machine learning can help people make a preliminary judgment about diabetes mellitus according to their daily physical examination data, and it can serve as a reference for doctors . For machine learning method, how to select the valid features and the correct classifier are the most important problems.

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. distinguished diabetes from normal people by using principal component analysis (PCA) and neuro fuzzy inference. used quantum particle swarm optimization (QPSO) algorithm and weighted least squares support vector machine (WLS-SVM) to predict type 2 diabetes proposed a system to predict diabetes, called LDA-MWSVM. In this system, the authors used Linear Discriminant Analysis (LDA) to reduce the dimensions and extract the features. In order to deal with the high dimensional datasets, built prediction models based on logistic regression for different onsets of type 2 diabetes prediction. focused on the glucose, and used support vector regression (SVR) to predict diabetes, which is as a multivariate regression problem. Moreover, more and more studies used ensemble methods to improve the accuracy . proposed a newly

ensemble approach, namely rotation forest, which combines 30 machine learning methods. proposed a machine learning method, which changed the SVM prediction rules.

Machine learning methods are widely used in predicting diabetes, and they get preferable results. Decision tree is one of popular machine learning methods in medical field, which has grateful classification power. Random forest generates many decision trees. Neural network is a recently popular machine learning method, which has a better performance in many aspects. So in this study, we used decision tree, random forest (RF) and neural network to predict the diabetes.

LITERATURE SURVEY

Existing problem :

A study was conducted to identify the prevalence of diabetes mellitus and the number of people affected with diabetes mellitus between the year 2000 and 2030. Prevalence of diabetes with regard to age and sex were estimated from 191 WHO states including US. In developing countries, people from urban area and rural area were separately taken for the study. Globally, people with diabetes mellitus were 2.8% in 2000 in all age groups. Whereas the prevalence diabetes mellitus is expected to increase at 4.4% by the year 2030 among all age groups. There were 171 million of patients with diabetes mellitus in 2000 and it is expected to raise the peak rate of 366 million by the year 2030. With regard to sex, men had higher prevalence of diabetes than women, but women with diabetes were higher than men. The results revealed that in developing countries, the urban people with diabetes were expected to double the number from 2000 to 2030. Further obese people had high prevalence of diabetes than non-obese people. So, obesity is one of the dangerous risk factors for diabetes mellitus

Proposed Solution :

Diabetic Mellitus Prediction using Auto AI project consist a Machine Learning model to predict the disease using

IBM Watson Auto AI Machine Learning Service. The model is deployed on IBM cloud to urge the value which might be used as API in mobile app or web app building. Here, we are going to be developing an internet application which is constructed using node red service.

We make use of the various user input values to the deployed model. The model prediction is then showcased on interface. This model is employed to predict the distribution based by using different input parameters.

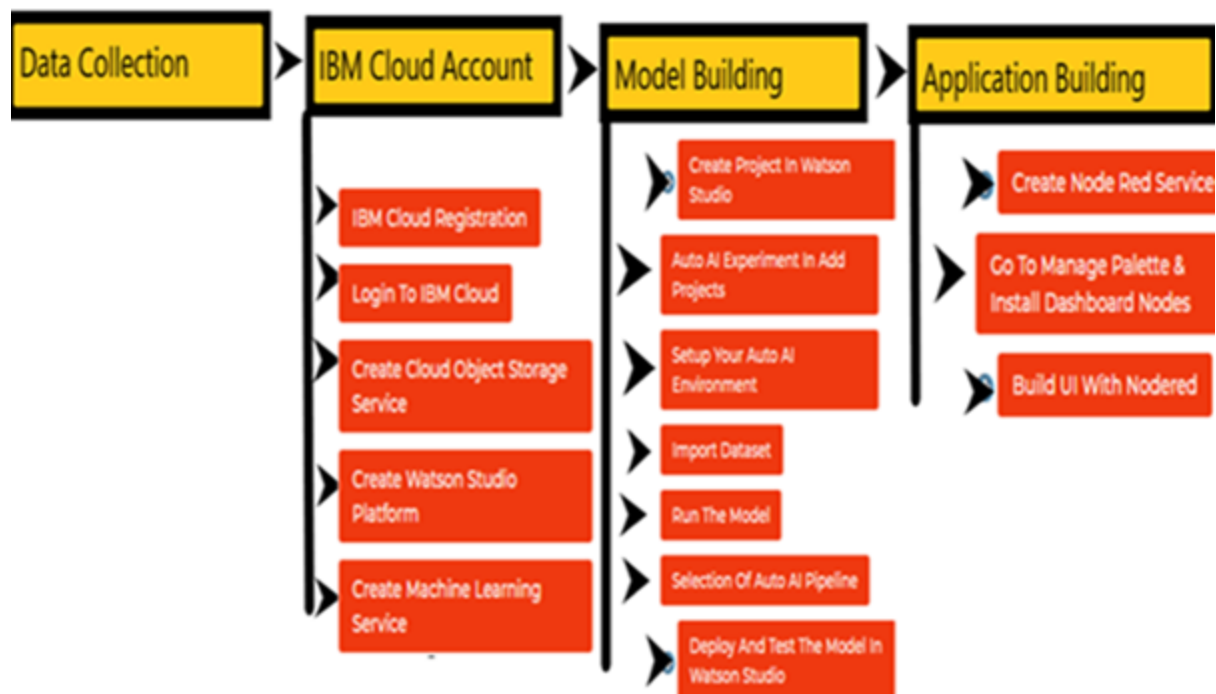
This thesis examines the appliance of machine learning algorithms to predict cost in effective manner. Machine learning techniques are often utilized in various areas to predict different values prediction. Such techniques would help companies to boost their services to customers.

EXPERIMENTAL INVESTIGATION

There are six steps in experimental investigation on of a general project:

1. Choose a Project Idea
2. Conduct Background Research
3. Compose a Hypothesis
4. Design your Experiment
5. Collect Data
6. Analyse Data and Draw Conclusions

All the data has been collected considering the above factors and it has been formatted. After formatting it has been uploaded in the project and after that using Watson Studio Auto AI Experiment, it is uploaded to cloud object storage service and implemented. Based on these implementation, the value can be predicted using the data we have collected. After that application is developed using Node red Service.



RESULT

After the implementation, deployment of project the result i.e. predicted diabetes mellitus can be seen in Node Red UI. This value depends on different parameters. The Node Red UI provide us simple way to get the result of Auto AI Experiment. Here is he Node Red UI predicts the Diabetes Mellitus

Default

text

plas *

148

pres *

72

skin *

35

test *

0

mass *

33.6

pedi *

0.627

age *

50

SUBMIT

CANCEL

ADVANTAGES

The advantages are easy to implement, accessibility is fast, continuous Improvement, wide application, available 24x7, no human intervention needed. We can handle multi-dimensional and multi-variety data.

DISADVANTAGES

The disadvantages are lack of security, loss of control on data, high error susceptibility, dependence of network/providers.

APPLICATION

Using The Auto AI Experiment, one can build and deploy a machine learning model with sophisticated training features. In the given project we can predict the price of the required vehicle by giving few input parameters.

CONCLUSION

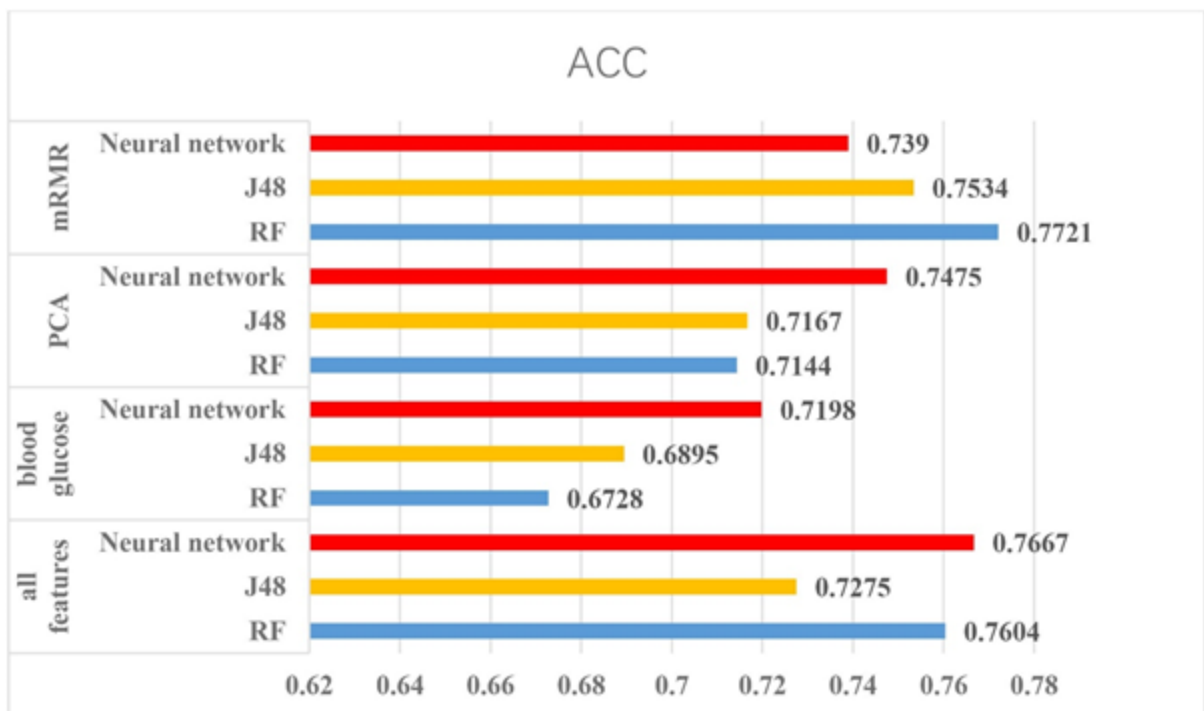
In this project by using IBM Cloud the model processing is been done in Auto AI services in IBM cloud and then the deployment is been done in Watson studio

and application is build using Node red service which has been successful as we are able to get the desire

FUTURE SCOPE

As we are developing day to day there is a continuous growth of Auto AI and Machine Learning. The web application can be used to predict the disease DIABETES MELLITUS , precisely and efficiently instead of n number of people being involved directly or indirectly.

The results of using Pima Indians dataset.



we can find PCA is not very suitable to the two dataset. And using all features has a good performance, especially for the Luzhou dataset. There is not much difference among random forest, decision tree and neural network when the feature set contains blood glucose. When we used the features without blood glucose, random forest has the best performance. But relatively speaking, the neural network performs poorly.

JSON FILE

```
[{"id":"1eb5ca72.88fc46","type":"tab","label":"Flow
1","disabled":false,"info":""},{ "id":"85db52f4.9253f","type":"function
","z":"1eb5ca72.88fc46","name":"PreToken","func":"global.set(\\\"ye\\\",ms
g.payload.ye)\\nvar
apikey=\\\"PFiputPROqo-ufpfzkjyNlYmEglDr7TZaT2eLqvJFp2v\\\";\\nmsg.headers=
{\\\"content-type\\\":\\\"application/x-www-form-urlencoded\\\"}\\nmsg.payload=
{\\\"grant_type\\\":\\\"urn:ibm:params:oauth:grant-type:apikey\\\",\\\"apikey\\\":
apikey}\\nreturn
msg;\",\"outputs\":1,\"noerr\":0,\"x\":255,\"y\":316.0000047683716,\"wires\":[[\"c
9e415db.50b328\"]]],{ \"id\":\"c9e415db.50b328\",\"type\":\"http
request\",\"z\":\"1eb5ca72.88fc46\",\"name\":\"\",\"method\":\"POST\",\"ret\":\"obj\",
\"paytoqs\":false,\"url\":\"https://iam.cloud.ibm.com/identity/token\",\"tls\":
\"\",\"proxy\":\"\",\"authType\":\"\",\"x\":454,\"y\":305,\"wires\":[[\"5110fcc4.37e3d4
\"]]],{ \"id\":\"ca6abf0c.e1515\",\"type\":\"inject\",\"z\":\"1eb5ca72.88fc46\",\"nam
e\":\"\",\"topic\":\"\",\"payload\":\"\",\"payloadType\":\"date\",\"repeat\":\"\",\"cronTa
b\":\"\",\"once\":false,\"onceDelay\":0.1,\"x\":100.5,\"y\":230.00000286102295,\"w
ires\":[[\"85db52f4.9253f\"]]],{ \"id\":\"fabe21f2.66578\",\"type\":\"debug\",\"z\":
\"1eb5ca72.88fc46\",\"name\":\"\",\"active\":true,\"tosidebar\":true,\"console\":f
```

```

also, "tostatus":false, "complete":"payload", "targetType":"msg", "x":893.
0000114440918, "y":47.00000190734863, "wires":[]}, {"id":"5110fcc4.37e3d4
", "type":"function", "z":"1eb5ca72.88fc46", "name":"Pre
Prediction", "func":"var ye = global.get('ye')\nvar
token=msg.payload.access_token\nvar
instance_id=\"a5c85af3-7889-44a1-8b37-b113804ffa5f\"\nmsg.headers={ 'Con
tent-Type': 'application/json', \"Authorization\": \"Bearer
\"+token, \"ML-Instance-ID\":instance_id}\nmsg.payload={\"fields\":
[\"years of experince\"], \"values\": [[ye]]}\nreturn
msg;\", \"outputs\":1, \"noerr\":0, \"x\":651.0000076293945, \"y\":268.000003814697
27, \"wires\":[[\"1f0edb6a.219855\"]]}, {\"id\":\"1f0edb6a.219855\", \"type\":\"http
request\", \"z\":\"1eb5ca72.88fc46\", \"name\": \"\", \"method\":\"POST\", \"ret\":\"obj\", \"
paytoqs\":false, \"url\":\"https://eu-gb.ml.cloud.ibm.com/v3/wml_instances/
a5c85af3-7889-44a1-8b37-b113804ffa5f/deployments/84ff180c-c180-4393-a034
-9272adacbc84/online\", \"tls\": \"\", \"proxy\": \"\", \"authType\": \"\", \"x\":744.500007
6293945, \"y\":212.00000381469727, \"wires\":[[\"c39682fe.a214c\"]]}, {\"id\":\"ee
561524.f0e538\", \"type\":\"ui_form\", \"z\":\"1eb5ca72.88fc46\", \"name\": \"\", \"label
\": \"\", \"group\":\"5a638ca3.b40164\", \"order\":0, \"width\":0, \"height\":0, \"options
\": [{\"label\":\"years of
experience\", \"value\":\"ye\", \"type\":\"number\", \"required\":true}], \"formValue\"
: {\"ye\":\"\", \"payload\":\"\", \"submit\":\"submit\", \"cancel\":\"cancel\", \"topic\":\"
\", \"x\":46, \"y\":359, \"wires\":[[\"85db52f4.9253f\"]]}, {\"id\":\"404f2519.c7877c\",
\"type\":\"ui_text\", \"z\":\"1eb5ca72.88fc46\", \"group\":\"5a638ca3.b40164\", \"orde
r\":1, \"width\":0, \"height\":0, \"name\": \"\", \"label\":\"text\", \"format\":\"{{msg.pay
load}}\", \"layout\":\"row-spread\", \"x\":772.6000366210938, \"y\":364.2000122070
3125, \"wires\":[]}, {\"id\":\"c39682fe.a214c\", \"type\":\"function\", \"z\":\"1eb5ca7
2.88fc46\", \"name\": \"\", \"func\":\"msg.payload=msg.payload.values[0][0]\nretu
rn
msg;\", \"outputs\":1, \"noerr\":0, \"x\":604.6000366210938, \"y\":107.600006103515
62, \"wires\":[[\"fabe21f2.66578\", \"404f2519.c7877c\"]]}, {\"id\":\"5a638ca3.b40
164\", \"type\":\"ui_group\", \"z\": \"\", \"name\":\"Default\", \"tab\":\"e10eba51.b60508\"
, \"disp\":true, \"width\":\"6\", \"collapse\":false}, {\"id\":\"e10eba51.b60508\", \"ty
pe\":\"ui_tab\", \"z\": \"\", \"name\":\"Home\", \"icon\":\"dashboard\", \"disabled\":false,
\"hidden\":false}

```

Node Red Web Application Link

<https://node-red-bwkfv.eu-gb.mybluemix.net/ui>

Conclusion

Diabetes mellitus is a disease, which can cause many complications. How to exactly predict and diagnose this disease by using machine learning is worthy studying. According to the all above experiments, we found the accuracy of using PCA is not good, and the results of using the all features and using mRMR have better results. The result, which only used fasting glucose, has a better performance especially in Luzhou dataset. It means that the fasting glucose is the most important index for predict, but only using fasting glucose cannot achieve the best result, so if want to predict accurately, we need more indexes. In addition, by comparing the results of three classifications, we can find there is not much difference among random forest, decision tree and neural network, but random forests are obviously better than the another classifiers in some methods. The best result for Luzhou dataset is 0.8084, and the best performance for Pima Indians is 0.7721, which can indicate machine learning can be used for prediction diabetes, but finding suitable attributes, classifier and data mining method are very important. Due to the data, we cannot predict the type of diabetes, so in future we aim to predicting type of diabetes and exploring the proportion of each indicator, which may improve the accuracy of predicting diabetes.

References

Alghamdi, M., Al-Mallah, M., Keteyian, S., Brawner, C., Ehrman, J., and Sakr, S. (2017). Predicting diabetes mellitus using SMOTE and ensemble machine learning approach: the henry ford exercise testing (FIT) project.

PLoS One 12:e0179805. doi: 10.1371/journal.pone.0179805

[PubMed Abstract](#) | [CrossRef Full Text](#) | [Google Scholar](#)

