iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** I **Month of publication:** January 2022

DOI: https://doi.org/10.22214/ijraset.2022.40090

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue I Jan 2022- Available at www.ijraset.com

Diabetes Prediction using BMI

Pushkar.V. Nerpagar[1], Prof. Vijaykumar Bhanuse[2]
[1]*Vishwakarma institute of technology*

Abstract: Technology interface with the health industry has revolutionised the industry in massive amount the surgeries that took large amount of time are now reduced to few minutes of work due to advancement in machinery same war the frontline process of the health is OPD is as important as it may look so diabetes is one of the rapid spreading human problem due to busy life format so this module work on preliminary basis

I. INTRODUCTION

Diabetes is not new to the human world and the are various test to predict the chances of getting it but using our simple test that can be measured at home using simple machines(sugar level indicator ,weight machine) these are some basic foods that we eat, specifically carbohydrate foods. Carbohydrate foods provide our body with its main energy source everybody, even those people with diabetes, needs carbohydrate. Carbohydrate foods include bread, cereal, pasta, rice, fruit, dairy products and vegetables (especially starchy vegetables). When we eat these foods, the body breaks them down into. Glucose. The glucose moves around the body in the bloodstream. Some of the glucose is taken to our brain to help us think clearly and function. The remainder of the glucose is taken to the cells of our body for energy and also to our liver, where it is stored as energy that is used later by the body. In order for the body to use glucose for energy, insulin is required. Insulin is a hormone that is produced by the beta cells in the pancreas. Insulin works like a key to a door. Insulin attaches itself to doors on the cell, opening the door to allow glucose to move from the blood stream, through the door, and into the cell. If the pancreas is not able to produce enough insulin (insulin deficiency) or if the body cannot use the insulin it produces (insulin resistance), glucose builds up in the bloodstream (hyperglycaemia) and diabetes develops. Diabetes Mellitus means high levels of sugar (glucose) in the blood stream and in the urine.

A. Symptoms of Diabetes

1) Frequent Urination
2) Increased thirst
3) Tired/Sleepiness
4) Weight loss
5) Blurred visio
6) Mood swings
7) Confusion and difficulty concentrating
8) Frequent infections

II. LITERATURE REVIEW

Uses the classification on diverse types of datasets that can be accomplished to decide if a person is diabetic or not. The diabetic patient's data set is established by gathering data from hospital warehouse which contains two hundred instances with nine attributes. These instances of this dataset are referring to two groups i.e., blood tests and urine tests. In this study the implementation can be done by using regressor to classify the data and the data is assessed by means of 10-fold cross validation approach, as it performs very well on small datasets, and the outcomes are compared. The logistics reg. is used. It was concluded that logistics works best showing an accuracy of 60.2% among others.

finds and calculate the accuracy, sensitivity and specificity percentage of numerous classification methods and also tried to compare and analyse the results of several classification methods in sklearn, the study compares the performance of same classifiers when implemented on some other tools which includes logistics and collab using the same parameters (i.e., accuracy, sensitivity and specificity). They applied logistics and decisonalgorithms. The result shows that the module has 76% of accuracy with logistic regression

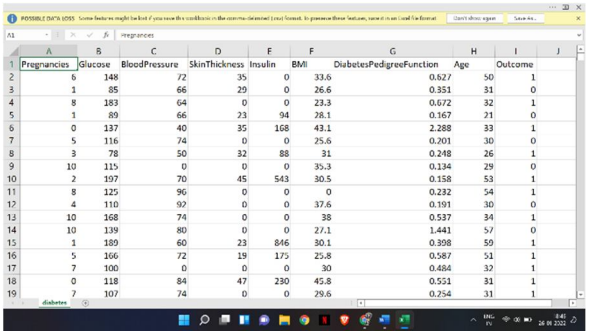
Considering the algorithm trained most important part its accuracy and how to increase the accuracy and make it more better and make it more efficiently checking the data shows it how clean the data and make it more null free and as the data becomes cleaner the accuracy of the algorithm also increases

©IJRASET: All Rights are Reserved

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue I Jan 2022- Available at www.ijraset.com

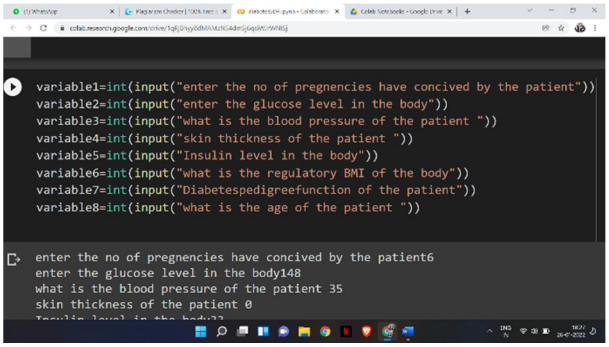
III. METHADOLOGY

Machine learning basically is like a mould it doesn't care about what you put in it just mould it's just trains whatever you put in it So, providing the clean data is the job of the data engineering after the data cleaning is done, we move to the programming part we initialise the library pandas and this library is basically used for getting or reading the data from the computer or whatever source we provide to it and data we provide for this project is of excel. After the data is received and read by the program now try to view the data and try to segmentize the data and sort the data for what our need of work from the data and segregate the data into two parts so the one part can be train with each other following is small representation of the data with its attributes:-



	A	B	C	D	F	F	G	H	I	J
1	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome	
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	

For this algorithm we have used various library and as program is understand the variable and implementation of it but when an interface is made so the program can be understand to its constraints so this program uses 7 variable assign to each 7 attribute that we took in with the data and ask 7 simple question about the attribute which make algorithm more user friendly following is illustration of the question asked: -



```
variable1=int(input("enter the no of pregnencies have concived by the patient"))
variable2=int(input("enter the glucose level in the body"))
variable3=int(input("what is the blood pressure of the patient "))
variable4=int(input("skin thickness of the patient "))
variable5=int(input("Insulin level in the body"))
variable6=int(input("what is the regulatory BMI of the body"))
variable7=int(input("Diabetespedigreefunction of the patient"))
variable8=int(input("what is the age of the patient "))
```

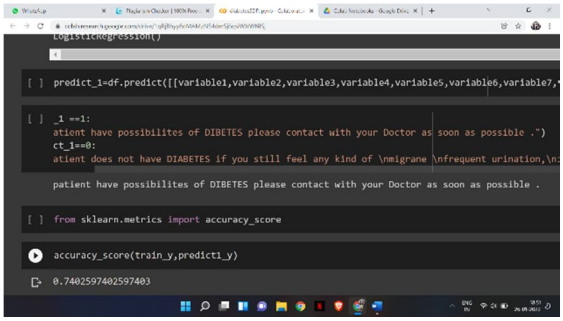
```
enter the no of pregnencies have concived by the patient6
enter the glucose level in the body148
what is the blood pressure of the patient 35
skin thickness of the patient 0
```

Whenever we create an deployable software for this algorithm this following will be easily asked question to get the real time data from the user and this increase the horoscope for the future use of the algorithm

Now initialise the sklearn library and import logistic regressor and fit the data and now just predict the data with provides data and initialise the variable and aske the data from the user and make the function actually working that's the end for the algorithm functioning

IV. RESULT AND DISCUSSION

For this algorithm the conclusion is user friendly interface and how this increases the future use spectrum for the project and also the successful of the logistic regression Following is the output of the algorithm when the patient is likely to have diabetes: -



```
LogisticRegression()

[ ] predict_1=df.predict([[variable1,variable2,variable3,variable4,variable5,variable6,variable7,

[ ] _1 ==1:
    atient have possibilites of DIBETES please contact with your Doctor as soon as possible .")
    ct_1==0:
    atient does not have DIABETES if you still feel any kind of \nmigrane \nfrequent urination,\n

    patient have possibilites of DIBETES please contact with your Doctor as soon as possible .

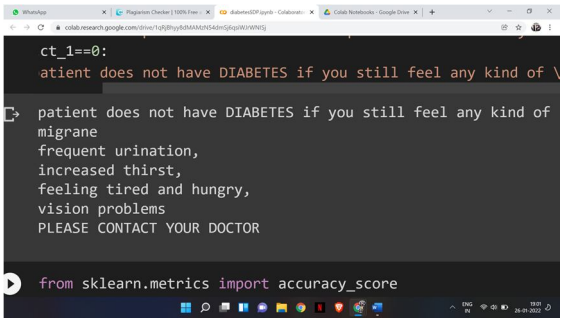
[ ] from sklearn.metrics import accuracy_score

    accuracy_score(train_y,predict1_y)

    0.7402597402597403
```

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

For patient with no possibilities of diabetes:-



In this way we can make this algorithm more interactive and make it more usable user point of view

V. HELPFULL HINTS

For making this programming programmer must have proper knowledge about python coding and should know about pandas and sklearn or basic machine learning algorithm notation and then we need data for training as for this algorithm we use the data provided for the diabetes

VI. LIMITATION

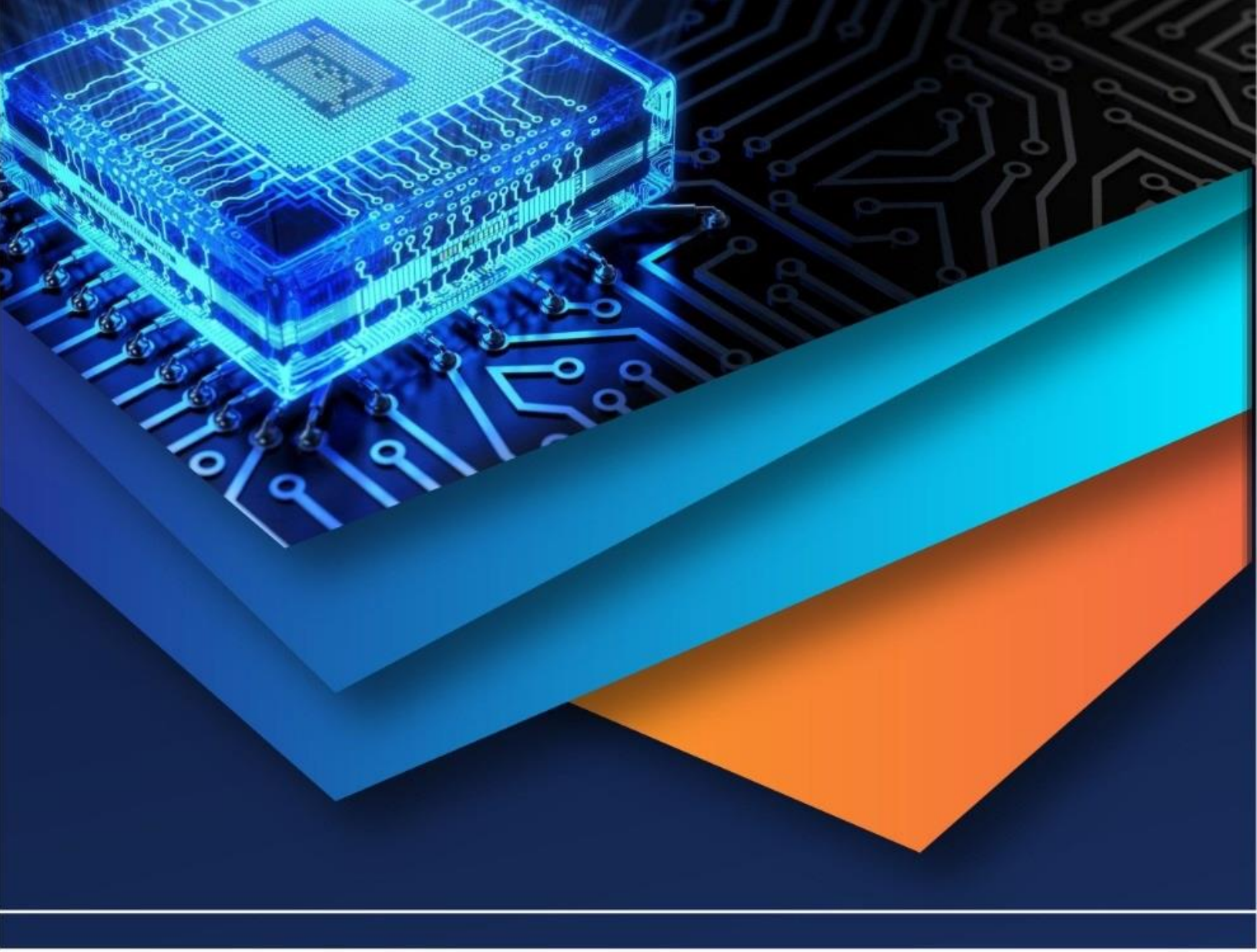
This is basic algorithm and the user info getting from the is programmer based for real time data is given by programmer and this algorithm will get its real use when used by a real life user or person suffering from this consequences

VII. CONCLUSION AND FUTURE WORK

One of the important real-world medical problems is the detection of diabetes at its early stage. In this study, systematic efforts are made in designing a system which results in the prediction of diabetes. During this work, five machine learning classification algorithms are studied and evaluated on various measures. Experiments are performed on john Diabetes Database. Experimental results determine the adequacy of the designed system with an achieved accuracy of 99% using Decision Tree algorithm. In future, the designed system with the used machine learning classification algorithms can be used to predict or diagnose other diseases. The work can be extended and improved for the automation of diabetes analysis including some other machine learning algorithms

VIII. ACKNOWLEDGEMENT

1) Department of CSE, IMS Engineering College, Ghaziabad, Uttar Pradesh, India
2) Online machine learning sources
3) Kaggle data provider



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)