AI-BASED DIABETES PREDICTIOION SYSTEM

1.INTRODUCTION

Diabetes is the fast growing disease among the people even among the youngsters. In understanding diabetes and how it develops, we need to understand what happens in the body without diabetes. Sugar (glucose) comes from the 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.

The glucose moves around the body in thebloodstream. Some of the glucose is taken to ourbrain to help us think clearly and function. Theremainder of the glucose is taken to the cells of ourbody for energy and also to our liver, where it isstored as energy that is required.

Diabetes is a chronic disease with the potential to cause a worldwide health care crisis. According to International Diabetes Federation 382 million people are living with diabetes across the whole world. By 2035, this will be doubled as 592 million.

Diabetes is a disease caused due to the increase level of blood glucose. This high blood glucose produces the symptoms of frequent urination, increased thirst, and increased hunger. Diabetes is a one of the leading cause of blindness, kidney failure, amputations, heart failure and stroke. When we eat, our body turns food into sugars, or glucose.

2.Types of Diabetes

Type 1 diabetes means that the immune system is compromised and the cells fail to produce insulin in sufficient amounts. There are no eloquent studies that prove the causes of type 1 diabetes and there are currently no known methods of prevention.

Type 2 diabetes means that the cells produce a low quantity of insulin or the body can't use the insulin correctly. This is the most common type of diabetes, thus affecting 90% of persons diagnosed with diabetes. It is caused by both genetic factors and the manner of living.

3. METHODOLOGY

In this section we shall learn about the various classifiers used in machine learning to predict diabetes. We shall also explain our proposed methodology to improve the accuracy. Five different methods were used in this paper. The different methods used are defined below. The output is the accuracy metrics of the machine learning models. Then, the model can be used in prediction

4.ACKNOWLEDGEMENT

I have completed this work under the mentorship of Dr. Pankaj Agarwal (Professor & Head) & Ms. Sapna Yadav (Assistant Professor), Department of Computer Science & Engineering at IMS Engineering College, Ghaziabad. I am doing a online summer internship on Machine Learning where I have learnt the various Machine Learning Algorithms from both of my mentors as Course Instructors. This work is been assigned as project assignments to us.

I would like to express my special thanks to both of my mentors for inspiring us to complete the work & write this paper. Without their active guidance, help, cooperation & encouragement, I would not have my headway in writing this paper. I am extremely thankful for their valuable guidance and support on completion of this paper.

Program

Input

```
df[['Glucose','BloodPressure','SkinThickness','Insulin','BMI
']] =
df[['Glucose','BloodPressure','SkinThickness','Insulin','BMI
0']].replace(0,np.NaN)
```

input

```
df.head()
```

Output

```
BloodPressure                35
SkinThickness               227
Insulin                     374
BMI                          11
DiabetesPedigreeFunction      0
Age                           0
Outcome                       0
dtype: int64
```

input

```
# The missing values will be filled with the median values
of each variable.
def median_target(var):
    temp = df[df[var].notnull()]
    temp = temp[[var,
'Outcome']].groupby(['Outcome'])[[var]].median().reset_index
()
    return temp
```

input

```
# The values to be given for incomplete observations are
given the median value of people who are not sick and the
median values of people who are sick.
columns = df.columns
columns = columns.drop("Outcome")
for i in columns:
    median_target(i)
    df.loc[(df['Outcome'] == 0 ) & (df[i].isnull()), i] =
median_target(i)[i][0]
    df.loc[(df['Outcome'] == 1 ) & (df[i].isnull()), i] =
median_target(i)[i][1]
```

input

```
# Missing values were filled.
df.isnull().sum()
```

output

```
BloodPressure                 0
SkinThickness                 0
```

```
Insulin                       0
BMI                           0
DiabetesPedigreeFunction      0
Age                           0
Outcome                       0
dtype: int64
```

input

```
# In the data set, there were asked whether there were any outlier observations compared to the 25% and 75% quarters.
# It was found to be an outlier observation.
for feature in df:
Q1 = df[feature].quantile(0.25)
    Q3 = df[feature].quantile(0.75)
    IQR = Q3-Q1
    lower = Q1- 1.5*IQR
    upper = Q3 + 1.5*IQR
if df[(df[feature] > upper)].any(axis=None):
        print(feature,"yes")
    else:
        print(feature, "no")
```

```
Pregnancies yes
Glucose no
BloodPressure yes
SkinThickness yes
Insulin yes
BMI yes
DiabetesPedigreeFunction yes
Age yes
Outcome no
```

Input

```
# We determine outliers between all variables with the LOF method
from sklearn.neighbors import LocalOutlierFactor
lof =LocalOutlierFactor(n_neighbors= 10)
lof.fit_predict(df)
```

Input

```
   #We choose the threshold value according to lof scores
threshold = np.sort(df_scores)[7] threshold
```

output

```
-1.740031580305444
```

Input

```
#We delete those that are higher than the threshold
outlier = df_scores > threshold
df = df[outlier]
```

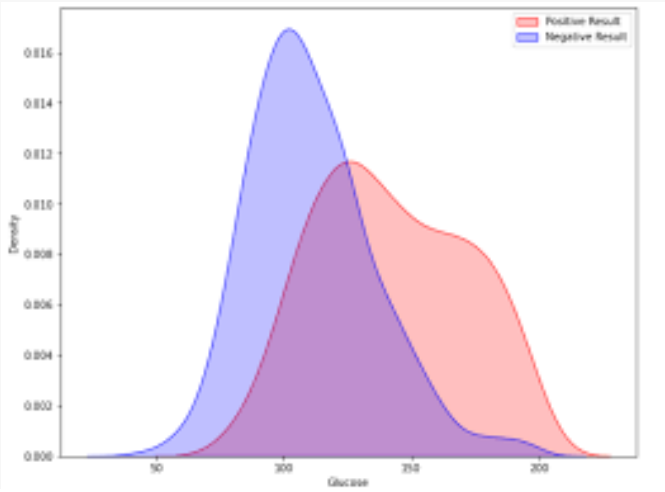
input

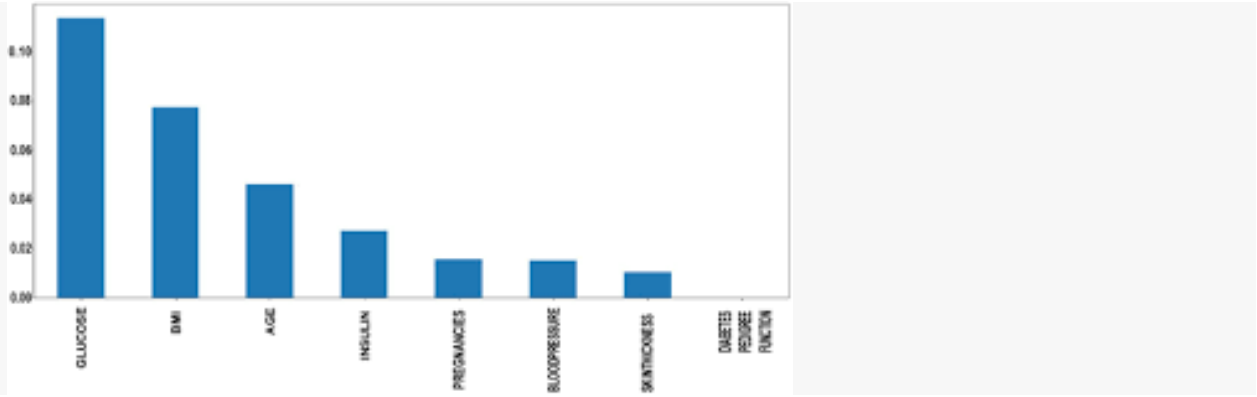
```
# The size of the data set was examined.
df.shape
```

output

```
(760, 9)
```

Bar chart





5.MACHINE LEARNING

These projects grounder in real world applications, ofter a comprehensive learning experience across diverse domains and technologies,enabling participants to bridges the theoretical pratical divide effectively.

Handling large amounts of data: With the ever-growing volume of data generated every day, it is increasingly difficult for humans to process and make sense of all this information. Machine learning can help businesses handle large amounts of data more efficiently and effectively and even use decision trees to take action on the information.

- **Reducing bias:** Machine learning algorithms are not biased toward certain data sets, unlike human beings, who may have personal biases that can distort their judgment. As a result, machine learning can help reduce bias in business decisions.
- **Improving accuracy:** Machine learning algorithms can achieve much higher accuracy than humans when making predictions or classifying labeled data. This improved accuracy can lead to better business outcomes and increased profits.
- **Discovering patterns and correlations:** Machine learning can help businesses uncover patterns and correlations in data that they may not have been able to detect otherwise. These learning systems can lead to better decision-making and a deeper understanding of the data.

6.conclusion

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 anachieved accuracy of 99% using Decision Tree algorithm.

In future, the designed system with the used machine learning classification algorithms can be extended and improved for the automation of diabetes analysis including some other machine learning algorithms.