Rameshwaram Institute of Technology and Management

Diabetes Prediction Using Machine Learning

Under The Supervision of

Er. Shubham Mishra

Group Member Name:

Prasuti Tiwari
Manisha Gautam

INTRODUCTION

Diabetes is a disease which is rapidly increasing all over the world. It occurs when pancreas does not produce sufficient insulin, or body can not sufficiently use insulin it produces. Diabetes person has increase blood glucose in the body. One of the major problem diabetic patients suffers from is the Diabetic Retinopathy (DR) and blindness. Since the number of diabetes patients is continuously increasing, it increases the data as well.

RESEARCH OUTCOME

- According to National Diabetes and Diabetic Retinopathy Survey 2019, one in 46 diabetic patients in India becomes visually impaired and one in 7 diabetic patients have vision impairment.

- This survey also states that between 2015 and 2019, around 16.9% of the Indian population suffered from DR.

- Prevalence of blindness among diabetic patient was 2.1% and visual impairment was 13.7%.

PROBLEM SPECIFICATION

- The most usual and conventional method for diagnosis and detection of diabetic retinopathy is by using human fundus images or retinal images.

- In our study, we focus on prediction of DR using health records of the diabetic patients.

- By using machine learning techniques, knowledge is acquired through these records, containing numerical values, to predict whether the patient is having DR or not.

- For this prediction of DR different classification algorithms (Support Vector Machine, K nearest neighbor, bagged trees, Logistic Regression) have been used.

CAUSES FOR DIABETES

- **What causes type 1 diabetes?**

- Type 1 diabetes occurs when your immune system, the body's system for fighting infection, attacks and destroys the insulin-producing beta cells of the pancreas. Scientists think type 1 diabetes is caused by genes and environmental factors, such as viruses, that might trigger the disease. Studies such as Trial Net External link are working to pinpoint causes of type 1 diabetes and possible ways to prevent or slow the disease.

WHAT CAUSES TYPE 2 DIABETES?

- Type 2 diabetes—the most common form of diabetes—is caused by several factors, including lifestyle factors and genes.
- **Overweight, obesity, and physical inactivity**
- **Insulin resistance**
- **Genes and family history**

EFFECTS OF DIABETES

- Diabetes can be effectively managed when diagnosed early. However, when left untreated, it can lead to potential complications that include:
- heart disease
- stroke
- kidney damage
- nerve damage

DATASET

	A	B	C	D	E	F	G	H	I
1	Pregnancie	Glucose	BloodPres	SkinThickn	Insulin	BMI	DiabetesP	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
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

	A	B	C	D	E	F	G	H	I
28	7	147	76	0	0	39.4	0.257	43	1
29	1	97	66	15	140	23.2	0.487	22	0
30	13	145	82	19	110	22.2	0.245	57	0
31	5	117	92	0	0	34.1	0.337	38	0
32	5	109	75	26	0	36	0.546	60	0
33	3	158	76	36	245	31.6	0.851	28	1
34	3	88	58	11	54	24.8	0.267	22	0
35	6	92	92	0	0	19.9	0.188	28	0
36	10	122	78	31	0	27.6	0.512	45	0
37	4	103	60	33	192	24	0.966	33	0
38	11	138	76	0	0	33.2	0.42	35	0
39	9	102	76	37	0	32.9	0.665	46	1
40	2	90	68	42	0	38.2	0.503	27	1
41	4	111	72	47	207	37.1	1.39	56	1
42	3	180	64	25	70	34	0.271	26	0
43	7	133	84	0	0	40.2	0.696	37	0
44	7	106	92	18	0	22.7	0.235	48	0
45	9	171	110	24	240	45.4	0.721	54	1
46	7	159	64	0	0	27.4	0.294	40	0
47	0	180	66	39	0	42	1.893	25	1
48	1	146	56	0	0	29.7	0.564	29	0
49	2	71	70	27	0	28	0.586	22	0
50	7	103	66	32	0	39.1	0.344	31	1
51	7	105	0	0	0	0	0.305	24	0
52	1	103	80	11	82	19.4	0.491	22	0
53	1	101	50	15	36	24.2	0.526	26	0
54	5	88	66	21	23	24.4	0.342	30	0

CODE AND OUTPUT:-

```
In [6]: df.describe()
```

Out[6]:

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
count	768.000000	768.000000	768.000000	768.000000	768.000000	768.000000	768.000000	768.000000	768.000000
mean	3.845052	120.894531	69.105469	20.536458	79.799479	31.992578	0.471876	33.240885	0.348958
std	3.369578	31.972618	19.355807	15.952218	115.244002	7.884160	0.331329	11.760232	0.476951
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.078000	21.000000	0.000000
25%	1.000000	99.000000	62.000000	0.000000	0.000000	27.300000	0.243750	24.000000	0.000000
50%	3.000000	117.000000	72.000000	23.000000	30.500000	32.000000	0.372500	29.000000	0.000000
75%	6.000000	140.250000	80.000000	32.000000	127.250000	36.600000	0.626250	41.000000	1.000000
max	17.000000	199.000000	122.000000	99.000000	846.000000	67.100000	2.420000	81.000000	1.000000

```
In [7]: df.shape
```

Out[7]: (768, 9)

```
In [8]: df['Outcome'].value_counts()
```

Out[8]: 0   500
        1   268
        Name: Outcome, dtype: int64

```
In [9]: df.groupby('Outcome').mean()
```

```python
In [9]: df.groupby('Outcome').mean()
```

Out[9]:

Outcome	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age
0	3.298000	109.980000	68.184000	19.664000	68.792000	30.304200	0.429734	31.190000
1	4.865672	141.257463	70.824627	22.164179	100.335821	35.142537	0.550500	37.067164

```python
In [10]: x=df.drop(columns='Outcome')
         y=df['Outcome']
```

```python
In [11]: x  #upto here is data preprocessing
```

Out[11]:

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age
0	6	148	72	35	0	33.6	0.627	50
1	1	85	66	29	0	26.6	0.351	31
2	8	183	64	0	0	23.3	0.672	32
3	1	89	66	23	94	28.1	0.167	21
4	0	137	40	35	168	43.1	2.288	33
...	...	...	...	...	...	...	...	...
763	10	101	76	48	180	32.9	0.171	63
764	2	122	70	27	0	36.8	0.340	27
765	5	121	72	23	112	26.2	0.245	30
766	1	126	60	0	0	30.1	0.349	47
767	1	93	70	31	0	30.4	0.315	23

```
In [13]: #Now we are standardizing The Data helps the machine learning model to get all the column values in range which helps in ma
         scaler=StandardScaler()
         scaler.fit(x)    #Fitting The Values of the data
         standard_values=scaler.transform(x) #Transforming it
         #PS We Can Still Use scaler.fit_transform
```

```
In [14]: x=standard_values
```

```
In [15]: x
```

```
Out[15]: array([[ 0.63994726,  0.84832379,  0.14964075, ...,  0.20401277,
                  0.46849198,  1.4259954 ],
                [-0.84488505, -1.12339636, -0.16054575, ..., -0.68442195,
                 -0.36506078, -0.19067191],
                [ 1.23388019,  1.94372388, -0.26394125, ..., -1.10325546,
                  0.60439732, -0.10558415],
                ...,
                [ 0.3429808 ,  0.00330087,  0.14964075, ..., -0.73518964,
                 -0.68519336, -0.27575966],
                [-0.84488505,  0.1597866 , -0.47073225, ..., -0.24020459,
                 -0.37110101,  1.17073215],
                [-0.84488505, -0.8730192 ,  0.04624525, ..., -0.20212881,
                 -0.47378505, -0.87137393]])
```

```
In [16]: #Train Test Split Model Selection
         x_train,x_test,y_train,y_test = train_test_split(x,y,test_size=0.3,stratify=y,random_state=2)
         #We Have Taken Four Variables x_train will train the data whereas x_test will test it
         #We don't want our ML model to know the data
         #Attributes of train test
         # 1. Giving All The Dataset 2.Giving The Test Size 3.Stratify helps to get exact proportions of diabetes and non diabetes
         #4.Random State just use to split the data
         #Stratify : This will be used by the train_test_split() function to ensure that both the train and test sets have the propa
```

PREVENTION

- Simple lifestyle measures have been shown to be effective in preventing or delaying the onset of type 2 diabetes. To help prevent type 2 diabetes and its complications, people should:
- achieve and maintain a healthy body weight;
- be physically active – doing at least 30 minutes of regular, moderate-intensity activity on most days. More activity is required for weight control;
- eat a healthy diet, avoiding sugar and saturated fats; and
- avoid tobacco use – smoking increases the risk of diabetes and cardiovascular disease.

CONCLUSION

- In this presentation machine learning methods are used for the prediction of Diabetic Retinopathy in patients, using their health records of diabetes.

- These health records were organized in a structured way by eliminating noisy data. With the help of machine learning algorithms, knowledge is extracted from these records in the form of numerical values for the prediction of DR.