AI model

Of

Diabetes Prediction

3rd semester project

Session 2019-2023

BS SOFTARE ENGINEERING

Department of Software Engineering

Faculty of Computer Science & Information Technology

The Superior College, Lahore

FALL 2023

Table of content

Contents

Table of figures

Artificial Intelligence: Prediction Model of diabetes diseases

Introduction:
This project aims to develop an AI-powered prediction model that can accurately forecast diabetic prediction based on historical data.

Modules:

Here are some modules that we have to import before going further.

```
import pandas as pd
import numpy as np
import pickle
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split
from sklearn import svm
from sklearn.metrics import accuracy_score
```

Figure 1

Methodology:
This project employed a machine learning approach to develop a loan status prediction model. The methodology consisted of the following steps:
1. Data Collection: A dataset containing loan applications, credit reports, and loan status (default or non-default) was collected from Kaggle.

Data loading

```
Ddata = pd.read_csv(filepath_or_buffer=r"D:\superior\semester_3rd\lab_AI\final project\diabetes.csv")
```

Figure 3

2. Data Preprocessing: The dataset was cleaned, transformed, and normalized to prepare it for modeling

Artificial Intelligence: Prediction Model of diabetes diseases

Data loading

```
1  Ddata = pd.read_csv(filepath_or_buffer=r"D:\superior\semester_3rd\lab_AI\final project\diabetes.csv")
```

Figure 4

```
1  Ddata.shape
```

```
(768, 9)
```

```
1  Ddata.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 768 entries, 0 to 767
Data columns (total 9 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   Pregnancies               768 non-null    int64
 1   Glucose                   768 non-null    int64
 2   BloodPressure             768 non-null    int64
 3   SkinThickness             768 non-null    int64
 4   Insulin                   768 non-null    int64
```

Figure 5

```
def conversion(cols):
    for col in cols:
        Ddata[col] = Ddata[col].astype("int64")
cols = ['BMI','DiabetesPedigreeFunction']
conversion(cols=cols)
```

```
Ddata.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 768 entries, 0 to 767
Data columns (total 9 columns):
```

Figure 6

```
Ddata.groupby(by='Outcome').mean()
```

```
Ddata.to_csv(path_or_buf=r"D:\superior\semester_3rd\lab_AI\final project\encoded_diabetes.csv")
```

Figure 7

3. Feature Engineering: Now we split data into x and y train and test for model evaluation.

Train Test split

+ Code + Markdown

```
x = Ddata.drop(labels=['Outcome'] , axis=1)
y = Ddata['Outcome']
```

Figure 8

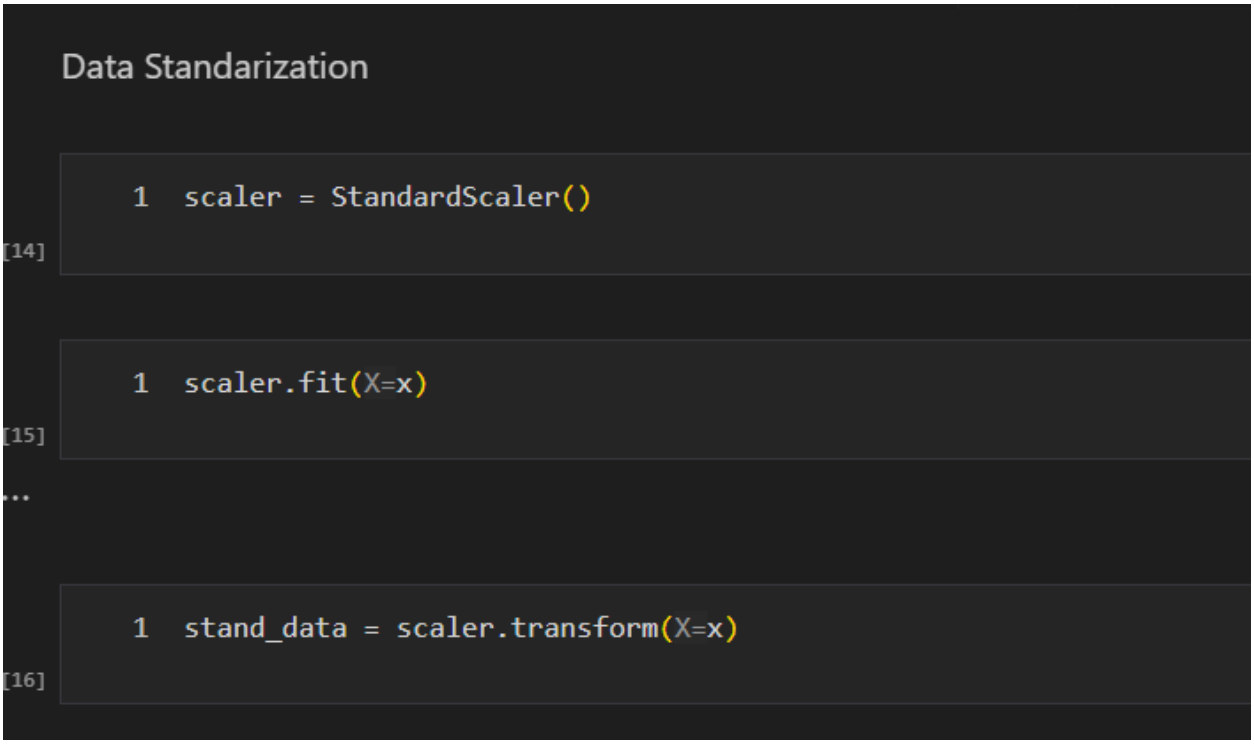

Figure 9

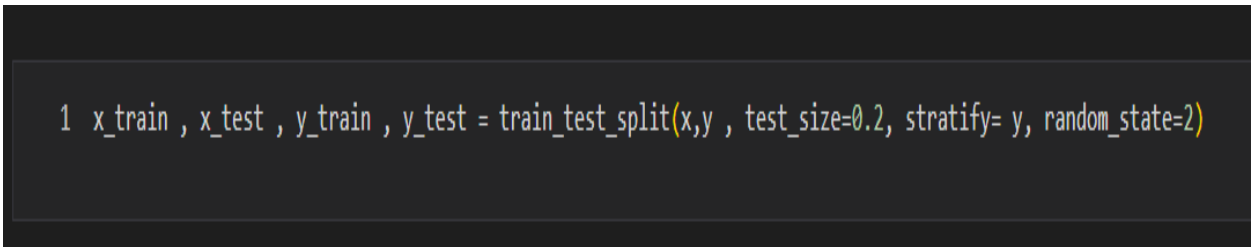

Figure 10

5. Model Evaluation: The performance of the selected model was evaluated using metrics such as accuracy, precision.

Training the model

```
[19]
1 model = svm.SVC(kernel='linear')
```

```
[20]
1 model.fit(X=x_train,y=y_train)
```

```
[21]
1 pickle.dump(obj=model,file=open(file='model_svc.pkl',mode='wb'))
```

Figure 11

Results:

The results of the project are presented below:

Model Performance: The random forest model outperformed other algorithms, achieving an accuracy of 80%,

```
1 train_accuracy = model.predict(X=x_train)
2 train_accuracy = accuracy_score(y_true=train_accuracy,y_pred=y_train)
```

```
1 print(f'Accuracy score of model: {train_accuracy}')
```

```
Accuracy score of model: 0.7736156351791531
```

```
1 test_accuracy = model.predict(X=x_test)
2 test_accuracy = accuracy_score(y_true=test_accuracy,y_pred=y_test)
```

Figure 12

Figure 13

Diabetic Prediction: The model accurately predicted loan status for 75% of the test dataset.

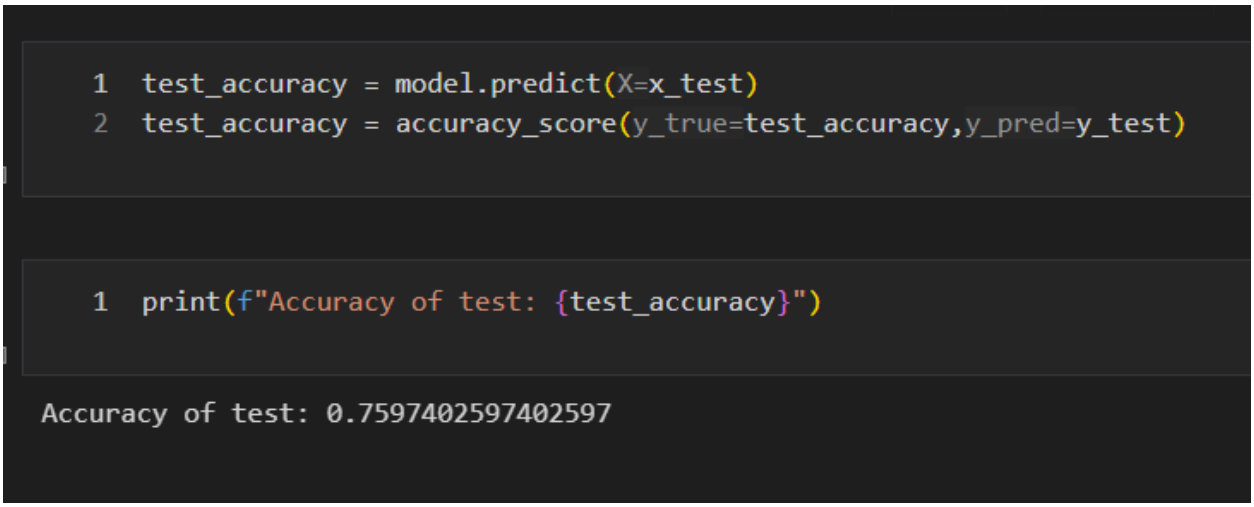

Figure 14