Main objective of the analysis

train classification model on the dataset to detect if the patient have diabeties or not

Brief description of the data set

The objective of the dataset is to diagnostically predict whether or not a patient has diabetes, The datasets consists of several medical predictor variables and one target variable, Outcome. Predictor variables includes the number of pregnancies the patient has had, their BMI, insulin level, age, and so on.It has a total of 768 rows and 9 columns

S No.	Column	Description	Data Type	Cate
1	Pregnancies	Number of times pregnant	Int	Discre
2	Glucose	Plasma glucose concentration a 2 hours in an oral glucose tolerance test	Int	Discre
3	Blood pressure	Diastolic blood pressure (mm Hg)	Int	Discre
4	Skin thickness	Triceps skin fold thickness (mm)	Int	Discre
5	Insulin	2-Hour serum insulin (mu U/ml)	Int	Discre
6	BMI	Body mass index (weight in kg/(height in m)^2)	Float	Contir
7	DiabetesPedigreeFunction	Diabetes pedigree function	Float	Contir
8	Age	Age (years)	Int	Discre
9	Outcome	Class variable (0 or 1)	Int	Discre

Plan for Data Exploration, Feature Engineering and Modelling

The steps in solving the Regression Problem are as follows:

1. Packages to be installed
2. Load the libraries
3. Load the dataset
4. General information about the dataset
5. Exploratory Data Analysis (EDA)
6. Modeling
7. Recommendations

Packages to be installed

1. autokeras

Show code

Load the libraries

1. numpy
2. pandas
3. matplotlib
4. seaborn
5. sklearn
6. autokeras
7. autopytorch
8. tqdm
9. tensorflow
10. pickle

Show code

Show code

Load the dataset

location of dataset

Show code

reading the dataset into dataframe

Show code

```
   0            6      148             72             35        0  33.6
   1            1       85             66             29        0  26.6
   2            8      183             64              0        0  23.3
   3            1       89             66             23       94  28.1
   4            0      137             40             35      168  43.1

      DiabetesPedigreeFunction  Age  Outcome
   0                     0.627   50        1
   1                     0.351   31        0
   2                     0.672   32        1
   3                     0.167   21        0
   4                     2.288   33        1
```

number of rows and coulmns in dataset

Show code

```
(768, 9)
```

dataset information

Show code

```
<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  
 5   BMI                       768 non-null    float64
 6   DiabetesPedigreeFunction  768 non-null    float64
 7   Age                       768 non-null    int64  
 8   Outcome                   768 non-null    int64  
```

Features Encoding

Show code

Split the data into test and train

Show code

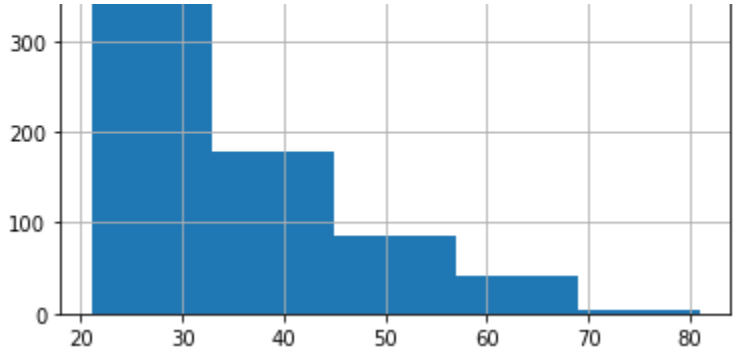
Show code

Exploratory Data Analysis (EDA)

Summary Statistics for Numerical columns

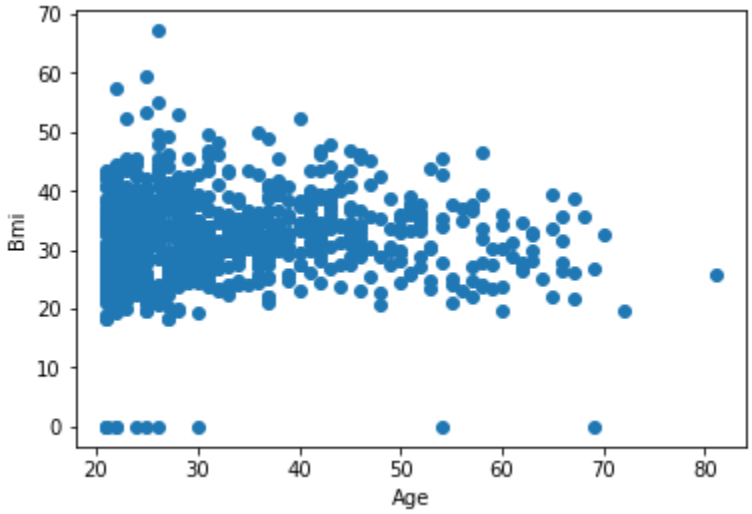
Show code

	Pregnancies	Glucose	Bloodpressure	Skinthickness	Insulin	\
count	768.000000	768.000000	768.000000	768.000000	768.000000	
mean	3.845052	120.894531	69.105469	20.536458	79.799479	
std	3.369578	31.972618	19.355807	15.952218	115.244002	
min	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	1.000000	99.000000	62.000000	0.000000	0.000000	

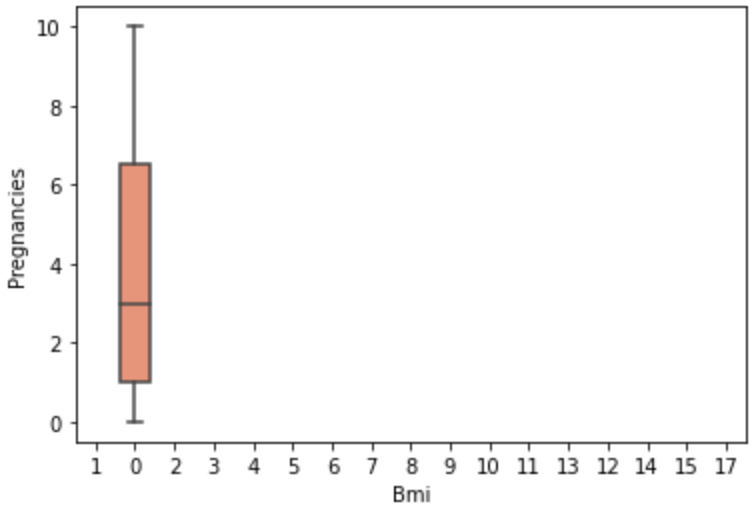


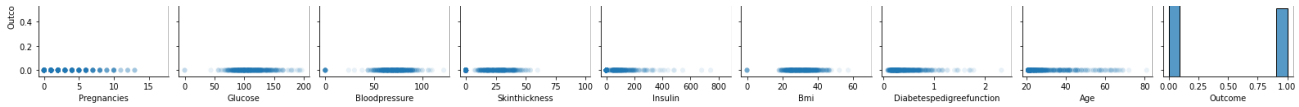
Correlation between Age and BMI

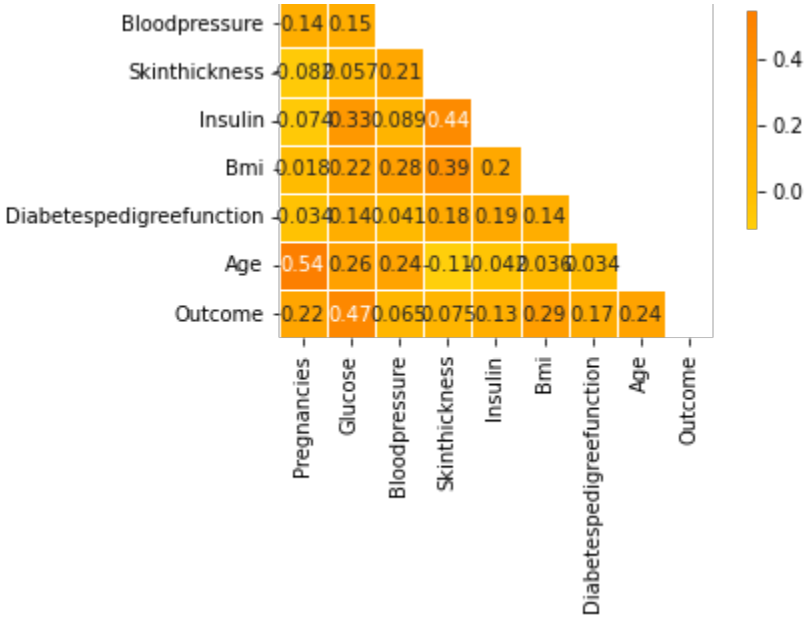
Show code



Show code




Outco
0.4
0.2
0.0
0
5
10
15
Pregnancies
0
50
100
150
200
Glucose
0
50
100
Bloodpressure
0
25
50
75
100
Skinthickness
0
200
400
600
800
Insulin
0
20
40
60
Bmi
0
1
2
Diabetespedigreefunction
20
40
60
80
Age
0.00
0.25
0.50
0.75
1.00
Outcome


Bloodpressure
Skinthickness
Insulin
Bmi
Diabetespedigreefunction
Age
Outcome
Pregnancies
Glucose
Bloodpressure
Skinthickness
Insulin
Bmi
Diabetespedigreefunction
Age
Outcome
0.4
0.2
0.0